UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

VOL. XXVI MARCH 19, 1929 NO. 29

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN No. 45

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A CRITICAL STUDY OF MEASURES OF ACHIEVEMENT RELATIVE TO CAPACITY

By

C. W. ODELL
Assistant Director, Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1929

The Bureau of Educational Research was established by act of the Board of Trustees June 1, 1918. It is the purpose of the Bureau to conduct original investigations in the field of education, to summarize and bring to the attention of school people the results of research elsewhere, and to be of service to the schools of the state in other ways.

The results of original investigations carried on by the Bureau of Educational Research are published in the form of bulletins. A list of available publications is given on the back cover of this bulletin. At the present time five or six original investigations are reported each year. The accounts of research conducted elsewhere and other communications to the school men of the state are published in the form of educational research circulars. From ten to fifteen of these are issued each year.

The Bureau is a department of the College of Education. Its immediate direction is vested in a Director, who is also an instructor in the College of Education. Under his supervision research is carried on by other members of the Bureau staff and also by graduates who are working on theses. From this point of view the Bureau of Educational Research is a research laboratory for the College of Education.

Bureau of Educational Research
College of Education
University of Illinois, Urbana

BULLETIN No. 45

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A CRITICAL STUDY OF MEASURES OF ACHIEVEMENT RELATIVE TO CAPACITY

By

C. W. Odell
Assistant Director, Bureau of Educational Research

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1929

3730 12 28 5584

UNIVERSITY
OF ILLINOIS
:: PRESS ::

TABLE OF CONTENTS

LIST OF TABLES

A CRITICAL STUDY OF MEASURES OF ACHIEVEMENT RELATIVE TO CAPACITY

CHAPTER I

DEFINITION OF TERMS AND STATEMENT OF PROBLEM

Introduction. The use of measures of achievement relative to capacity has been one of the most enthusiastically recommended and widely employed procedures that have arisen in connection with the standardized test movement. Almost at once after the first specific public suggestions as to how to compute such measures, they were received with popular favor by both the leaders and the rank and file of the educational profession. As is frequently true, so in this instance, the majority of those who adopted such procedures did so non-critically.[1] From time to time some unusually thoughtful worker called attention to the very serious deficiencies of the measures employed, but only recently has a considerable amount of attention been focused upon this point. It has, therefore, appeared worth while to devote this bulletin to a critical study of measures of achievement relative to capacity. Before proceeding further, a number of more or less technical terms employed in the discussion will be defined, so that there may be no doubt as to their exact significance. Three of these are rather general; the others are names of various measures of achievement and intelligence.

Achievement. The expression "achievement" is used to refer to the quantity and quality of school work done by pupils. Thus a measure of achievement is a measure of how much school work pupils have covered and how well they have mastered it.

Capacity. As employed in this bulletin, "capacity" is limited to mental capacity or potential ability to do school work. It is most often determined by the use of general intelligence tests. That is, a pupil's capacity is considered to be indicated by the score he makes upon such a test.

General intelligence (or intelligence). As implied in the preceding paragraph, "general intelligence" or merely "intelligence" is used as synonymous with capacity to do school work. In other words, a

[1]As an example of the non-critical attitude toward the use of measures of achievement relative to capacity, the following quotation from an article which appeared in 1922 may be given. "We believe that the Accomplishment Quotient is the fairest and most valuable measure of both the efficiency of teacher and the pupil, that by relying on it for guidance, the teacher will come to exact from him that hath even more than he has been given and take from him that hath not even less than he has been able to give."

pupil's general intelligence is thought of as his potential ability to learn the subject-matter presented in school.

Classified list of measures to be defined. The measures of achievement and intelligence to be defined may be classified as follows, first as absolute[2] or relative[3] and then under further subdivisions:

I. Absolute measures[4]
   A. Measures of intelligence
      1. Mental age
   B. Measures of achievement
      1. Achievement age
      2. Accomplishment age
      3. Attainment age
      4. Subject age
      5. Educational age
   C. Measures of either intelligence or achievement
      1. T-score[5] (or sigma score)

II. Relative measures
   A. Measures of intelligence relative to chronological age[6]
      1. Quotient measures[7]
         a. Intelligence quotient
      2. Sigma measures
         a. Mental index[8]

---

[2]Absolute measures are those that express achievement or intelligence, as the case may be, directly in terms of some unit of measurement of the characteristic measured and not by comparison with some other characteristic.

[3]Relative measures are those which compare absolute measures of one characteristic with those of another and express the first relative to the second.

[4]No attempt will be made to give definitions of all suggested absolute measures of either capacity or achievement. Only those that have to date been employed in securing relative measures, and therefore are referred to in this bulletin, will be defined. All of the absolute measures defined except the T-score are age measures. It is a sigma measure; that is, it is expressed in terms of the sigma or standard deviation of the scores of the whole group. The standard deviation is that distance which, if laid off on either side from the mean of a normal distribution, includes 34.13 per cent of all the cases.

[5]The T-score is most commonly employed as a measure of achievement, but may equally as well be used for intelligence.

[6]A number of other relative measures in addition to the two named above have been proposed, but as none of them are involved in the computation of achievement relative to capacity, they will not be defined. The references describing more fully the relative measures are not given here, but may be found in Chapter II in connection with the accounts of their origins.

[7]The term "quotient measure" is employed here in a general sense to refer to any measure that involves the division of one quantity by another. It would be equally appropriate to use the expression "ratio measure."

[8]The reader may wonder why the mental index, and later the educational index, are classed as relative measures rather than as absolute measures, since they as well as the T-score are sigma measures. The reason is that they are based upon distributions for each age and thus really show capacity or achievement, as the case may be, relative to the age of the individual concerned, whereas the T-score is based upon a single age group and does not indicate capacity or achievement relative to the individual's age.

B. Measures of achievement relative to chronological age
   1. Quotient measures
      a. Subject quotient
      b. Educational quotient
      c. Accomplishment quotient (sometimes)
   2. Sigma measures
      a. Educational index[9]

C. Measures of achievement relative to capacity
   1. Quotient measures
      a. Achievement quotient
      b. Accomplishment quotient
      c. Attainment quotient
      d. Achievement ratio
      e. Accomplishment ratio
      f. Subject ratio
      g. Educational ratio
      h. Efficiency quotient
   2. Difference measures[10]
      a. Pintner's difference
      b. Symond's index of studiousness or of effort

Mental age (M.A.). Mental age is ordinarily used to refer to a score on a general intelligence test expressed in terms of age units. A mental age of a given amount is equal to the average intelligence of an unselected or random group of pupils of that age. For example, if the average score made by pupils eleven years and six months of age upon a certain general intelligence test is 95, that score is said to be equivalent to a mental age of eleven years and six months.

Achievement age (A.A.). Achievement age is similar to mental age except that it represents a pupil's score upon an achievement rather than upon a general intelligence test. Therefore, an achievement age of a given amount, such as ten years and eight months, is a means of expressing the average score made by pupils of that chronological age.

Accomplishment age (A.A.). This expression is often used instead of achievement age. The two are entirely synonymous.

Attainment age (A.A.). This is synonymous with achievement age and accomplishment age, but is rarely used.

---

[9]*See* note on mental index.
[10]A difference measure is one found by subtracting one measure from another.

Subject age (S.A.). Subject age is somewhat synonymous with achievement age, but differs in that it is limited to a pupil's age score in a single subject. Thus a pupil may have a subject age in arithmetic, a subject age in reading, and so on.

Educational age (E.A.). Educational age is likewise largely synonymous with achievement age. The difference is that it is applied only to a pupil's average standing in a number of school subjects, whereas achievement age may refer either to such an average or to his standing in a single subject.

T-score (or sigma score). A T-score is one given according to the T-scale, which is based upon the distribution of ability of an unselected group of twelve-year-old pupils. The scale consists of 100 units of .1 standard deviation each, extending from five standard deviations below average twelve-year-old ability to five standard deviations above average. T-scores, therefore, range from 0 to 100, with 50 as average. They are not commonly used in the case of pupils whose ages differ from twelve by more than three or four years, but very similar scores may be found for groups of such pupils. All scores of this general sort are often called sigma scores.

Intelligence quotient (I.Q.). The intelligence quotient is found by dividing an individual's mental age by his chronological age. In formula form, $\text{I.Q.} = \frac{\text{M.A.}}{\text{C.A.}}$. In writing the result, it is carried to two places and the decimal point omitted. Thus an individual whose mental age is the same as his chronological age, or, in other words, the same as the average mental age of all persons of his chronological age, has an intelligence quotient of 100. If his mental age is greater than his chronological age, his intelligence quotient is proportionately greater, and if less, it is less. For persons in the upper teens or above, the actual chronological age is not employed as the divisor, but instead sixteen or some other fixed age supposed to represent the point at which the growth of intelligence ceases.

Mental index (M.I.). This measure performs the same function as the intelligence quotient in that it compares the intelligence of an individual with the average intelligence of persons of his age, but the method of computation is decidedly different. It is determined according to a scale based upon an assumption of normal distribution of ability of pupils of each age, and ranges from 0 through 50 as average or normal up to 100. Its chief difference from the T-score is that it is based upon the distribution of scores of the age group

to which the individual belongs rather than always upon that of the same age group.

Subject quotient (S.Q.). There is the same difference between the subject quotient and the second or less usual meaning of the achievement quotient as between the subject age and the achievement age. That is, the subject quotient refers to results from a single subject alone, whereas the achievement quotient may also apply to a combination of results from several subjects. It is found by the formula $S.Q. = \frac{S.A.}{C.A.}$.

Educational quotient (E.Q.). This is nearly synonymous with achievement quotient in its second or less common sense, since $E.Q. = \frac{E.A.}{C.A.}$. It differs in the same way as does the educational age from the achievement age, in that it is employed only to refer to combined or average results in several school subjects, whereas the latter may also refer to a single subject.

Educational index (E.I.). The educational index is similar to the mental index except that it is derived from scores on achievement rather than intelligence tests.

Achievement quotient (A.Q.). Two chief methods of securing an achievement quotient have been suggested. One of these involves the division of the achievement age by the mental age; that is, $A.Q. = \frac{A.A.}{M.A.}$.[11] This is the more usual of the two and may be considered the standard method; therefore, unless otherwise stated, the writer will employ the term in this sense. The less common method of securing an achievement quotient is to divide achievement age by chronological age. It is used in this sense by those who employ achievement ratio instead of achievement quotient for achievement age divided by mental age. As is true of the intelligence quotient, so the achievement and all other quotients and ratios are regularly carried to two places and written without a decimal point. The expression has also been used by several persons with meanings varying somewhat from either of those just given, but these minor differences will not be dealt with here.

Accomplishment quotient (A.Q.). This term is synonymous with achievement quotient.

---

[11]Sometimes the formula $A.Q. = \frac{E.Q.}{I.Q.}$ is employed. It yields exactly the same result as the one given in the text. *See* p. 13f.

Attainment quotient (A.Q.). This is synonymous with achievement quotient and accomplishment quotient, but very rarely employed.

Achievement ratio (A.R.). Achievement ratio is synonymous with achievement quotient in its first or more frequent sense. That is, $A.R. = \frac{A.A.}{M.A.}$.

Accomplishment ratio (A.R.). This is synonymous with achievement ratio.

Subject Ratio (S.R.). This is similar to the achievement ratio except that it is limited to a single subject. $S.R. = \frac{S.A.}{M.A.}$ or $\frac{S.Q.}{I.Q.}$.

Educational ratio (E.R.). This is similar to the achievement ratio except that it always refers to combined results from several school subjects. In other words, $E.R. = \frac{E.A.}{M.A.}$ or $\frac{E.Q.}{I.Q.}$.

Efficiency quotient (Eff.Q.). This is a suggested but very rarely used measure of achievement relative to capacity. It is secured by dividing a pupil's achievement point score by the norm for his age and then further dividing this result by his intelligence quotient.

Pintner's difference. This is a measure little used except by Pintner and those associated with him. It is found by subtracting an individual's mental index from his educational index. Thus a positive difference indicates that a pupil is doing better work than is done by average pupils of his capacity, a difference of zero that he is doing work of the same quality, and a negative difference that he is doing work of an inferior quality.

Index of studiousness or of effort. This is a rather general term proposed by Symonds to include various more or less similar methods of comparing achievement with capacity. These methods are especially intended for use in high school, where the difficulties connected with employing the various quotient and ratio measures are serious. Two possible methods of computing it are suggested. One of them consists merely in ranking pupils according to their achievement and also according to their capacity, and taking the differences. The second is somewhat more difficult. It requires that both achievement and general intelligence test scores be turned into standard deviation units. The difference between the two standard deviation scores of each pupil is then found, multiplied by ten, and added algebraically to fifty.[12]

[12]The multiplication by ten is merely to avoid fractional indices, and the addition to fifty to eliminate negative ones.

Purpose and plan of this study. It has already been stated that the general purpose of this bulletin is to present a critical study of measures of achievement relative to capacity. As preliminary to this study, a brief historical account of the origin of various proposed measures of the sort mentioned will be given. This will be found in Chapter II. Following that, an attempt will be made to answer three chief questions, as follows:

1. What are the merits and demerits of the various proposed measures?
2. How valid[13] are such measures?
3. How reliable[14] are such measures?

Chapter III will deal with the first of these questions, Chapter IV with the second, and Chapter V with the third. As part of his attempt to give answers, the writer will refer to practically all published critical discussions pertaining to the topic and will also present some hitherto unpublished data which he has gathered. Chapter VI will contain a brief summary and the general conclusions. Certain supplementary data and discussions will be given in Appendices A, B, and C.

---

[13]A test or measure is valid or possesses validity if it fulfills the function which it is intended or stated to perform. It may lack validity either because it is unreliable or because it measures some other ability or abilities than the statement of its function specifies. Since few, if any, tests possess perfect validity, the term is used relatively, and tests are said to be valid when they approach perfect validity.

[14]A test or measure is reliable, or has reliability, if a second application of the test yields scores equivalent to those obtained on the first application or correlating perfectly with them. In other words, a test is perfectly reliable not only if each pupil makes exactly the same score the second time as the first, but also if there is a constant and known difference, either an amount or a ratio, between the first set of scores and the second. For example, if each pupil's score were four points greater the second time than the first, or if it were increased by 10 per cent of his original score, reliability would be said to be perfect. Since even the best standardized tests are not perfectly reliable, the term is used in a relative sense to refer to those which approach perfect reliability.

CHAPTER II

A BRIEF ACCOUNT OF PROPOSED MEASURES

The educational and subject quotients. Apparently the first quotient or ratio scores to receive use in connection with standardized tests were the educational and subject quotients. These, along with the educational and subject ages, were suggested by McCall and employed by him and his students before 1920. It appears, however, that these measures were not mentioned in print until later than the achievement age and the achievement quotient. As may be seen by referring to the definitions of the educational and subject quotients given in Chapter I, they are measures similar to the previously well accepted intelligence quotient[1] in that they compare pupil performance with chronological age.

The achievement quotient. There is some doubt as to who should receive most credit in connection with the idea of comparing achievement test scores with those on intelligence tests by the quotient or ratio method. The use of standardized achievement tests, which began in 1908, became fairly common within a few years, so that by 1915 they were being used in many school surveys and by many administrators, supervisors, and teachers. At first, practically all persons who interpreted the results of such tests appear to have done so without regard to the capacity of the pupils with whom they were dealing. Occasionally someone more thoughtful than most made an attempt to interpret achievement in the light of pupils' intelligence, but it was not until about 1920 that this procedure began to be discussed in print, in public addresses, or elsewhere.

Among the earliest workers to point out the desirability of comparing pupils' achievement with their intelligence, or capacity to achieve, were the Presseys. In an address at the seventh Indiana University Conference on Educational Measurements in April, 1920, Mrs. Pressey devoted considerable attention to this point.[2] Indeed, she even made use of the expression "quotient of achievement" in connection with the comparison of achievement and capacity. She did not, however, employ this phrase in the same delimited sense in which "achievement quotient" soon came to be used, but in a much

---

[1]For the origin of the intelligence quotient, *see*:
Freeman, F. N. *Mental Tests.* Boston: Houghton Mifflin Company, 1926, p. 98.

[2]Pressey, L. C. "The Relation of Intelligence to Achievement in the Second Grade," *Bulletin of the Extension Division,* Vol. 6, No. 1. Bloomington, Indiana: Indiana University, September, 1920, p. 68-77.

more general way to refer to any measure which served the purpose indicated. Furthermore, although the Presseys continued to emphasize the importance of making such comparisons, they did not press the use of this or any similar term.

Franzen thought of employing a quotient or ratio score at least as early as 1919 when he was a student under McCall and appears to deserve credit for originating the idea. However, he did not publish it immediately, and the first proposal of the achievement quotient as now employed to appear in print seems to have been in connection with the Illinois Examination published by the Bureau of Educational Research of the University of Illinois, in 1920.[3] In response to a demand for a battery of tests which could be employed for general survey purposes, the Monroe Standardized Silent Reading Tests, Revised, the Monroe General Survey Scale in Arithmetic, and an intelligence test later called the Illinois General Intelligence Scale, were combined. In connection with this battery, the already rather common practice of transmuting scores made upon an intelligence scale into mental ages was followed. In addition, Monroe suggested that scores upon the reading and arithmetic tests be turned into achievement ages. He also provided for the computation of achievement quotients[4]—that is, achievement ages divided by mental ages—as measures of achievement compared with capacity to achieve.

The accomplishment quotient. As has already been stated, Franzen had previously conceived the idea of a measure similar to that suggested by Monroe, but did not publish it until later. This measure, which he termed the "accomplishment quotient," appears to have been first mentioned in print in the fall of 1920,[5] a few months after Monroe's proposal had appeared. He suggested a different way of computing it which, however, differs from Monroe's only in method and not in result. Franzen's method is to divide the educational quotient by the intelligence quotient. Since $\text{E.Q.} = \frac{\text{E.A.}}{\text{C.A.}}$ and $\text{I.Q.} = \frac{\text{M.A.}}{\text{C.A.}}$, it can readily be seen that the formula $\text{A.Q.} = \frac{\text{E.Q.}}{\text{I.Q.}}$

[3]Monroe, W. S. and Buckingham, B. R. "Illinois Examination: Teacher's Handbook." Urbana: University of Illinois, Bureau of Educational Research, 1920. 31 p.

Monroe, W. S. and Buckingham, B. R. "The Illinois Examination I and II: Teacher's Handbook." Bloomington, Illinois: Public School Publishing Company, 1920. 32 p.

Monroe, W. S. "The Illinois Examination," *University of Illinois Bulletin,* Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. 70 p.

[4]At the time Monroe conceived the idea of the achievement quotient, he had not yet heard of Franzen's similar idea.

[5]Franzen, R. H. "The Accomplishment Quotient of School Marks in Terms of Individual Capacity," *Teachers College Record,* 21:432-40, November, 1920.

easily reduces to $A.Q. = \frac{E.A.^6}{M.A.}$, which is synonymous with $\frac{A.A.}{M.A.}$.

The accomplishment ratio. Although the use of the achievement or accomplishment quotient at once began to be very common, Franzen appears to have been dissatisfied with the latter term. Less than two years later he began to advocate the use of a different expression, "accomplishment ratio," for the same idea.[7] At this time he employed the term "quotient" to refer to a measure secured when the divisor is chronological age and "ratio" to refer to one secured when the divisor is mental age. Moreover, he practically dropped accomplishment quotient in favor of subject quotient and educational quotient.[8] Furthermore, he employed accomplishment ratio in a more limited sense than he had previously used accomplishment quotient, limiting it to educational age divided by mental age, and employing the expression "subject ratio" for subject age divided by mental age. Although many workers have followed Franzen in his use of accomplishment ratio instead of accomplishment quotient, the tendency has been not to limit it to educational age divided by mental age, but rather to use it for either that or subject age divided by mental age. In other words, when used it has generally been entirely synonymous with achievement quotient as suggested by Monroe and with Franzen's first use of accomplishment quotient.

Although it was undoubtedly desirable that distinct terms be employed to refer to the relation of achievement to chronological age and to mental age or capacity, and although there appears to have been no logical objection to the terms and distinctions suggested by Franzen, they did not come into general use. Monroe's achievement quotient and Franzen's similar accomplishment quotient, both abbreviated A.Q., had become broadly enough known and used by the date of Franzen's later suggestions that the latter did not cause a general change of practice in respect to the matter. On the other hand, quite a number of persons followed Franzen in his later suggestions. Thus, although his first proposal was more generally accepted than the second, confusion was introduced in that accomplishment quotient was sometimes used in the one sense and sometimes in the other. To complicate the situation still more, it even occasionally happened that

---

[6]All that is necessary to make this reduction is to cancel the common denominator C.A.

[7]Franzen, R. H. "The Accomplishment Ratio," *Teachers College, Columbia University Contributions to Education,* No. 125. New York: Bureau of Publications, Columbia University, 1922. 59 p.

Franzen, R. H. "The Conservation of Talent," Terman, L. M., *et al. Intelligence Tests and School Reorganization.* Yonkers: World Book Company, 1922, Chapter IV.

[8]Although at this time (1922) Franzen dropped the term "accomplishment quotient," he later employed it as synonymous with educational quotient or subject quotient.

those who used achievement or accomplishment quotient in the sense suggested by Monroe, and by Franzen in 1920, employed accomplishment ratio for accomplishment age divided by chronological age. In other words, they used A.Q. and A.R. in just the reverse senses from those which Franzen advocated in 1922. In the meantime the educational and subject quotients continued to receive more use, but, however, with no difference of opinion or practice as to their meaning, both being used to refer to achievement as related to actual age; that is, to $\frac{\text{E.A.}}{\text{C.A.}}$ and $\frac{\text{S.A.}}{\text{C.A.}}$, respectively.

Pintner's difference method. Before the quotient or ratio technique had become thoroughly established, another method of comparing achievement with capacity was suggested. In connection with his general survey and mental tests, Pintner proposed the use of a mental index, an educational index, and finally the difference between the two.[9] This appears to have been the first proposal for the use of a difference between achievement and intelligence test scores rather than the use of a quotient or ratio between them as a measure of achievement relative to capacity. Probably because the concept of mental age and the accompanying intelligence quotient were widely understood and firmly established in educational usage, and because of the similarity thereto of age and quotient or ratio measures of school achievement as contrasted with Pintner's indices and difference, the latter were never received into the same popular favor as the former. Few persons except Pintner and his students and co-workers made use of them and they were very rarely if ever employed in connection with other tests than his.

The efficiency quotient. Another suggested method of comparing achievement with capacity was that of Torgerson, who called his measure the "efficiency quotient."[10] As preliminary to securing it, he defined the achievement quotient in a new way, somewhat synonymous with the meaning of the subject or educational quotient. Instead of the common method of transmuting a pupil's achievement score into a subject age and then dividing by the chronological age, he suggested that the achievement score be divided by the norm or average score for the pupil's age, using both in terms of point scores.[11] The

---

[9]Pintner, Rudolf and Marshall, Helen. "A Combined Mental-Educational Survey," *Journal of Educational Psychology,* 12:32-48, January, 1921.

Pintner, Rudolf. "Manual of Directions for the Non-Language Mental and Educational Survey Tests." Columbus, Ohio: College Book Store, 1920. 16 p.

[10]Torgerson, T. L. "The Efficiency Quotient as a Measure of Achievement," *Journal of Educational Research,* 6:25-32, June, 1922.

[11]A point score is the score yielded directly by a test in terms of exercises done correctly, level of difficulty reached, or otherwise.

result, which he called the achievement quotient, is divided by the intelligence quotient to yield the efficiency quotient. Thus, in a general way, the latter has a similar significance to Monroe's achievement quotient and Franzen's 1920 accomplishment quotient, in that it represents a comparison of achievement with capacity. Torgerson justified his method by stating that a grade norm represented the performance of the average pupil, or in other words, was based upon an I.Q. of 100.

High-school and college accomplishment quotients. a. Peters' proposal. From time to time someone has pointed out that none of the commonly employed methods of computing quotients or ratios is adequate in the case of school subjects for which achievement ages can not be satisfactorily determined. This is a condition that holds for practically, if not absolutely, all high-school and college subjects. Achievement therein depends to a relatively small degree upon age, and much less upon capacity regardless of time spent upon the subject than in elementary school. Also differences in high-school curricula are so great that pupils of decidedly varying ages and mental abilities may be pursuing the same portion of the same subject. A suggestion for taking care of the situation has been offered by Peters,[12] who gave an empirical formula developed from a study of the marks of a considerable number of students at Ohio Wesleyan University. At first he tried a method that may be described as follows: The academic marks of students and also their intelligence scores were transmuted into T-scores[13]—that is, into standard deviation units—and then accomplishment quotients secured by dividing academic T-scores by intelligence T-scores. As this method was applied, however, it was found to have a serious defect in that persons ranking highest or lowest in intelligence could not by any possible means secure accomplishment quotients higher or lower, respectively, than 1.00,[14] since the highest academic T-score just equalled the highest intelligence T-score and similarly for the lowest of each. The same injustice was involved in the case of all students to the extent to which they stood near either end of the distribution of intelligence scores.

To correct this injustice, the empirical formula already referred to was developed. It was based upon two considerations: first, that

[12]Peters, C. C. "A Method for Computing Accomplishment Quotients on the High-School and College Levels," *Journal of Educational Research*, 14:99-111, September, 1926.

[13]For a fuller explanation and discussion of T-scores than is given on p. 8, *see*:
McCall, W. A. *How to Measure in Education*. New York: The Macmillan Company, 1922, p. 272 f., or:
Monroe, W. S. *An Introduction to the Theory of Educational Measurements*. Boston: Houghton Mifflin Company, 1923, p. 150 f.

[14]Peters does not follow the usual practice of omitting the decimal point in accomplishment quotients, but retains it. Therefore, his quotient of 1.00 is equivalent to a quotient of 100 in ordinary usage, of .85 to 85, and so on.

the correction should be an addition for students above average and a deduction for those below average sufficient to offset the amount which each is above or below the mean; second, that since students of high intelligence who exceed average accomplishment may be expected to do as much better than the average, and likewise those of inferior intelligence who fall below it to do as much worse, as students of normal or average intelligence, the addition or subtraction need only be made to the extent to which the normal chances have been exhausted. The average deviation from the median for the middle quintile[15] was found to be $.13\sigma$;[16] therefore, those of highest intelligence should have a chance to earn $.13\sigma$ on the average above normal, and those of lowest intelligence should on the average fall that much below, in their achievement quotients. The midpoint of the distribution should be 5, since the zero point of a T-scale or other scale based on standard deviation units is commonly taken at -5σ and the upper end of the scale at $+5\sigma$, so that its total range is 10σ and, of course, the midpoint 5σ above the zero point. In view of these considerations, the formula which Peters suggested to satisfy the conditions was as follows, using somewhat different terminology from that which he employed:

$$A.Q. = \frac{A}{I} \pm \frac{.13}{(5 \pm 4 - I) \quad (5 \pm 4 - A)}.$$

In this formula, A equals the academic mark expressed in terms of T-scores or standard deviation units, and I the intelligence test score similarly expressed. Wherever the plus or minus sign occurs in the formula, plus is to be used if the student's intelligence score is greater than 5σ and minus if it is less than 5σ. Peters suggested further that the formula not be employed unless both intelligence and academic accomplishment are above 6σ or below 4σ, since between these limits the correction is not more than .01 and therefore not worth the trouble of computing and applying.

b. Otis' suggestion. In his article, Peters also stated certain objections to his formula advanced by Otis and likewise a substitute proposed by the latter. This substitute is that the academic mark expressed in terms of the standard deviation be divided by the intelligence score in the same terms after the latter has been multiplied by the coefficient of correlation between the two, adjusted until the cor-

[15]Quintile is synonymous with fifth. Thus the middle quintile of any group includes the fifth of the individuals above and below which there are two-fifths when all are arranged in order according to the trait or measurement in question.

[16]The small Greek letter sigma, σ, is the commonly used abbreviation for the standard deviation.

relation between I.Q.'s and A.Q.'s is made zero. Otis' reasoning is stated as follows:

"The line of the means of the academic point-averages regresses from the line of relation so that at any point the latter falls only r_{AI} times as far above or below the mean as the former. For any given position in the intelligence series we can find the mean academic achievement that corresponds to it by multiplying the intelligence quotient by r_{AI}. Since this is the mean *A*, academic attainment, made by those with this *I*, we may take it as the 'normal' or 'standard' *A* for this *I*. The accomplishment quotient would then be the attained *A* divided by the derived one, that is,

$$\text{A.Q.} = \frac{\text{A}}{r_{AI}}$$

where each is measured from the mean. Starting from -5σ as a zero point,

$$\text{A.Q.} = \frac{5 + \text{A}}{5 + \text{I} \cdot r_{AI}}.\text{''}^{17}$$

Symonds' index of studiousness or of effort. Another proposal for comparing the achievements with the capacities of secondary-school pupils has been made by Symonds.[18] He accepted the achievement ratio as satisfactorily filling the need for such a measure in the elementary school, but showed, as have others also, that decidedly serious difficulties were connected with its use in high school. Therefore he proposed the use of a measure which he called the "index of studiousness" or "index of effort." As may be seen by looking at its definition in Chapter I,[19] Symonds used these terms in a general way and not as being limited to a single method of computing such a measure. In fact, he gave two possible methods of doing so, and implied that others as well might be used. Though differing in detail, both of his methods are based upon differences rather than quotients. He prefers the second, the one based upon standard deviation scores, to the first, even though it is more difficult. In his discussion, however, he admitted that both possess certain common weaknesses. Among these are that they show a regression effect and that they do not permit comparison between members of different classes, but only between those of a single class.

Nygaard's method of computing accomplishment quotients. Another proposal of a different method of computing the accomplishment quotient has been made by Nygaard.[20] His suggestion, however, was not intended to take care of the situation in high schools or other places where age norms are not available, but rather to

[17]Peters, C. C. "A Method for Computing Accomplishment Quotients on the High-School and College Levels," *Journal of Educational Research*, 14:99-111, September, 1926.

[18]Symonds, P. M. *Measurement in Secondary Education*. New York: The Macmillan Company, 1927, p. 521-25.

[19]*See* p. 10.

[20]Nygaard, P. H. "A Revised Accomplishment Quotient," *Journal of Educational Research*, 18:87, June, 1928.

remedy what he considered a defect in the usual method of computing A.Q.'s. It has been shown by various studies that it is almost impossible for a child with a high I.Q. to earn a high A.Q. and likewise that a child with a low I.Q. rarely has a low A.Q. Nygaard proposed to correct this condition. He defined accomplishment quotient as the educational or achievement age divided by the predicted educational or achievement age instead of by the mental age. In other words, $\text{A.Q.} = \dfrac{\text{A.A.}}{\text{Predicted A.A.}}$. The predicted A.A. is to be determined from mental age by the ordinary regression equation[21] as follows:

$$\text{Predicted A.A.} = r\frac{\sigma_{\text{A.A.}}}{\sigma_{\text{M.A.}}}(\text{M.A.} - M_{\text{M.A.}}) + M_{\text{A.A.}}.$$

That is, the predicted achievement age is found by adding to the mean achievement age ($M_{\text{A.A}}$) the product of the coefficient of correlation[22] (r) of achievement age with mental age times the standard deviation of the first ($\sigma_{\text{A.A.}}$) over that of the second ($\sigma_{\text{M.A.}}$) times the difference between mental age (M.A.) and mean mental age ($M_{\text{M.A.}}$).

Nygaard stated that by his method the average A.Q. for any group will be 100 irrespective of how the group ranks in achievement, the negative correlation between A.Q. and I.Q. will be eliminated, allowance will be made for the common fact that a group's achievement age tends to have a smaller range of variability than does its mental age, and any handicaps common to a group because of inferior instruction or other causes will not operate to cause lower A.Q.'s.

Rand's suggestion. Following rather severe criticisms of the statistical validity of measures of achievement relative to capacity, Miss Rand[23] has suggested what she calls a program of reconstruction of such measures. Since, however, her suggestions have to do entirely with the matter of statistical validity, they will not be considered at this point, but rather in Chapter IV, which deals with that topic.

Summary. A year or two previous to 1920, McCall and his students began to employ the educational and subject quotients. The achievement quotient was first mentioned in print by Monroe in 1920, although Franzen at least a year earlier had conceived the same idea

[21]A regression equation is a rectilinear or first-degree equation showing the relationship between two series of paired measures so that if one of a pair is known, the best possible prediction as to the magnitude of the other can be made.

[22]The coefficient of correlation, abbreviated r, is a measure of the rectilinear or straight-line relationship existing between two sets of paired measures. It ranges in value from +1.00, denoting perfect positive or direct relationship, through .00, which indicates that there is purely chance relationship, to −1.00, denoting perfect negative or inverse relationship.

[23]Rand, Gertrude. "A Discussion of the Quotient Method of Specifying Test Results," *Journal of Educational Psychology,* 16:599-618, December, 1925.

and employed orally the synonymous term "accomplishment quotient." In 1922, Franzen suggested "accomplishment ratio" instead of "accomplishment quotient," employing the latter for another purpose. About the same time Pintner suggested his difference method, but this did not meet with a great degree of acceptance. The same is true of the "efficiency quotient" suggested by Torgerson. Peters and Otis have proposed methods for computing high-school and college accomplishment quotients, but neither appears to have received any use. In 1927, Symonds suggested an "index of effort," which may be computed in any one of several ways. Still more recently, Nygaard has suggested a change in the method of computing accomplishment quotients. Present practice may be summarized by saying that the achievement quotient suggested by Monroe, and its synonym, the original accomplishment quotient of Franzen, are widely used, the subject quotient, educational quotient, and the accomplishment ratio somewhat less so, and the others rarely or not at all.

CHAPTER III

MERITS AND DEMERITS OF THE VARIOUS PROPOSED MEASURES

Problem of this chapter. As was stated on page 11, the problem of this chapter is to present the merits and demerits of the various proposed measures of achievement relative to capacity. This general problem, however, may be broken up into a number of parts, which deal with the following questions:

1. What is the most desirable terminology in cases where two or more terms with identical meanings have been proposed?
2. What are the advantages and disadvantages of comparing achievement with chronological age as contrasted with comparing it with mental capacity?
3. What significance should be attached to an achievement quotient or similar measure of 100?
4. What are the advantages and disadvantages of quotient measures as contrasted with difference measures?
5. What are the merits and demerits of each of the following suggested measures?
   a. Torgerson's efficiency quotient
   b. Peters' accomplishment quotient
   c. Otis' accomplishment quotient
   d. Symonds' index of studiousness or of effort
   e. Nygaard's accomplishment quotient

Before proceeding to the consideration of these questions, one important limitation of the discussion in this chapter should be noted. Since Chapter IV is devoted to the question of statistical validity and Chapter V to that of reliability, these topics will not be included in the general discussion of this chapter, although some reference will be made to reliability in answering the second and fourth questions stated above. On the whole, however, the merits and demerits of various proposed measures will be dealt with without regard to their validity and reliability.

Terminology. As was shown in Chapter I, there has been an unnecessary and confusing multiplication of terms and also considerable use of the same terms with different meanings. Because the term "quotient" has become much more widely used than "ratio" in comparing achievement with capacity, the writer recommends that it alone

be employed and the latter dropped entirely. Furthermore, the expression "attainment quotient" has been so very rarely used that it also may be eliminated for the sake of reducing unnecessary duplication. This leaves "achievement quotient" and "accomplishment quotient" to be applied to the achievement or accomplishment age divided by the mental age, and the writer urges that these terms, and these only, be used with this meaning. Theoretically, it might be still more desirable to employ just one of them and drop the other, but as both have come into such general use, and as they are very similar, it seems well to retain and approve both. With regard to the variations in the usual achievement quotient proposed by Peters, Otis, Nygaard and perhaps others, the writer recommends that these measures be referred to as Peters' accomplishment quotient, Otis' accomplishment quotient, Nygaard's accomplishment quotient, and so forth. For the result obtained by dividing by chronological instead of mental age, the expressions "subject quotient" and "educational quotient" may well be retained, the former to refer to a quotient based upon performance in a single school subject and the latter to one based upon average performance in several.

The question of differences in terminology has not arisen in connection with difference measures except that Symonds has suggested that either "index of studiousness" or "index of effort" be used for the measures which he suggests. It does not seem to the writer that there is any weighty reason why either one of these terms should be preferred to the other, but it does seem desirable that if this measure be employed, a single and universally used name be given it. He ventures to suggest, therefore, that the second term, "index of effort," be the preferred one. This recommendation is made for at least two reasons. The word "effort" is probably somewhat more commonly used and understood than "studiousness" and also is shorter.

Mental age versus chronological age as the standard of comparison. The comparison of achievement with capacity to achieve as expressed in terms of general intelligence or mental age is in most cases much more significant than its comparison with chronological age or the mere number of years an individual has lived. In other words, as Sherrod[1] and others have pointed out, the achievement quotient is ordinarily a decidedly more significant measure of relative achievement than is the subject or educational quotient. Within the last few years, a considerable mass of evidence has been accumulated

[1]Sherrod, C. C. "The Development of the Idea of Quotients in Education," *Peabody Journal of Education,* 1:44-49, July, 1923.

which indicates that what is called general intelligence is a considerably more potent factor in determining pupil achievement than is chronological age or even the number of years spent in school. That is, an individual's mental ability contributes more largely to his performance in the school subjects than does the length of time for which he has been exposed to the chance to learn.[2] Therefore, for practically all purposes for which relative measures of achievement are employed, it is more helpful to compare achievement with intelligence. It is ordinarily perfectly legitimate to expect a pupil to do school work corresponding to his intelligence, but not to his chronological age unless he happens to possess average intelligence for that age. Moreover, for these reasons, it is fair to rate teachers according to how well they capitalize their pupils' mental abilities, but not according to how great achievements they secure from pupils of a given chronological age.

Ruch,[3] among others, has suggested that since the correlation between achievement age and mental age in a single grade is practically always positive, whereas that between the achievement age and chronological age in a single grade is negative, the former comparison is better. This reason alone does not seem sufficiently strong to justify the use of $\frac{\text{A.A.}}{\text{M.A.}}$ rather than $\frac{\text{A.A.}}{\text{C.A.}}$, but is evidence for the fact stated above, that achievement depends upon mental ability rather than mere age.

An argument sometimes advanced in favor of the use of the educational quotient and subject quotient instead of the achievement quotient is that the former correlate rather highly with the intelligence quotient, whereas the latter practically always exhibits a negative correlation therewith.[4] Beeson and Tope,[5] for example, cite data which show a correlation of .90 between the E.Q. and the I.Q., whereas that

[2]The statement made above is true in general, although there may be exceptions to it in individual cases. No one doubts, for example, that a child of relatively low intelligence who has been in school for several years, will do better in the school subjects than one of high intelligence who has never, either at school, at home, or elsewhere, had the opportunity of reading, spelling, working with numbers, and so forth. Such extreme cases, however, are almost non-existent, as are also those even very nearly approaching them.

Evidence to support the statement made in the text may be found in the following references:

Heilman, J. D. "The Relative Influence upon Educational Achievement of Some Hereditary and Environmental Factors," *Twenty-Seventh Yearbook of the National Society for the Study of Education,* Part II. Bloomington, Illinois: Public School Publishing Company, 1928, p. 35-65.

Denworth, K. M. "The Effect of Length of School Attendance upon Mental and Educational Ages," *Twenty-Seventh Yearbook of the National Society for the Study of Education,* Part II. Bloomington, Illinois: Public School Publishing Company, 1928, p. 67-91.

[3]Ruch, G. M. "The Achievement Quotient Technique," *Journal of Educational Psychology,* 14:334-43, September, 1923.

[4]*See* Appendix A for a brief discussion of how such negative correlation may be reduced.

[5]Beeson, M. F. and Tope, R. E. "The Educational and Accomplishment Quotients as an Aid in the Classification of Pupils," *Journal of Educational Research,* 9:281-92, April, 1924.

between the A.Q. and the I.Q. for the same cases is —.46. Many other writers, among whom may be named Murdoch,[6] MacPhail,[7] Toops and Symonds,[8] Popenoe,[9] Wilson,[10] and Franzen[11] likewise report data showing negative correlations of considerable size between achievement quotients and intelligence quotients. It is not apparent to the writer, however, why this fact should be an argument against employing the achievement quotient. The usefulness of the A.Q. does not appear to depend at all on whether it correlates positively or negatively with the I.Q. Indeed, the very fact that a negative correlation is practically always found has called to our attention a significant condition which needs remedial attention. It is true that some educators and others had realized that most instruction in our schools was better adapted to average and dull pupils than to bright ones, but the finding of many negative correlations of the sort just referred to has brought the fact home in such a pointed fashion as to arouse a much keener and more general realization of the need.

Another argument occasionally urged against the use of mental age in preference to chronological age as the divisor is that the resulting quotient is more unreliable. This is, of course, true. The A.Q. is based upon two unreliable quantities, A.A. and M.A., whereas the S.Q. or E.Q. is based upon only one, since chronological age can ordinarily be determined with a high degree of accuracy. This argument does not seem, however, to possess any considerable validity. The mere fact that one measure is more reliable than another does not justify its use when it possesses little significance. If the same argument were carried further, it would do away with the use of achievement age in the numerator, since it also is unreliable, and would require that some measure which can be determined with perfect or almost perfect reliability be substituted. Since we possess no, or practically no, such measure of school achievement, measurement activity along that line would have to cease.

Significance of an A.Q. of 100. In his discussion of the accomplishment ratio, Franzen[12] appeared to regard an accomplishment

[6]Murdoch, Katharine. "The Accomplishment Quotient—Finding and Using It," *Teachers College Record,* 23:229-39, May, 1922.

[7]MacPhail, A. H. "The Correlation Between the I.Q. and the A.Q.," *School and Society,* 16:586-88, November 18, 1922.

[8]Toops, H. A. and Symonds, P. M. "What Shall We Expect of the A.Q.?" *Journal of Educational Psychology,* 13:513-28, December, 1922; 14:27-38, January, 1923.

[9]Popenoe, Herbert. "A Report of Certain Significant Deficiencies of the Accomplishment Quotient," *Journal of Educational Research,* 16:40-47, June, 1927.

[10]Wilson, W. R. "The Misleading Accomplishment Quotient," *Journal of Educational Research,* 17:1-10, January, 1928.

[11]Franzen, Raymond. "The Accomplishment Ratio," *Teachers College, Columbia University Contributions to Education,* No. 125. New York: Bureau of Publications, Teachers College, Columbia University, 1922. 59 p.

[12]*Ibid.*

ratio, or, as more commonly expressed, an achievement quotient, of 100 as the maximum if errors were eliminated. He argued that if the E.Q., which equals $\frac{\text{E.A.}}{\text{C.A.}}$, is ever greater than the I.Q., which equals $\frac{\text{M.A.}}{\text{C.A.}}$, it is probably because of spurious differences[13] and that, therefore, A.Q.'s secured by the formula $\text{A.Q.} = \frac{\text{E.Q.}}{\text{I.Q.}}$ will not exceed 100 except for the presence of accidental errors. As a number of writers, among whom are Toops and Symonds,[14] Symonds,[15] and Foran,[16] have pointed out, this concept is entirely untenable. There is no theoretical limit above which the A.Q. may not rise, although in actual practice it is rarely greater than 200, and in only a small per cent of cases exceeds 150.[17] Since an age norm, whether mental or achievement, of a given amount is the average performance of an unselected group of pupils of that age, it necessarily follows that the average A.Q. of an unselected group must be 100. For Franzen's concept to be valid it would be necessary to assume that no pupil's achievement could rise above the average of pupils of his mental age. If none could rise above the average it would follow that none could fall below it, and therefore that all would be just the same. However, we know that because some pupils study harder and by better methods than others, are more interested in their subjects, receive more outside help, and so on, they do better work than average pupils of their intelligence, and that the reverse of these causes is the reason why others do worse work than the average. Franzen's idea of some measure which would show how well pupils are achieving in comparison with the best they can do rather than with what average pupils actually do achieve, possesses some value, but this comparison cannot be made by means of any of the achievement quotients. Some of the proposed difference methods can be adapted to serve this purpose.

An argument sometimes advanced against the use of the achievement quotient or any similar measure with a mean value of 100 is

---

[13]The differences are spurious because they are due to the unreliability of the test or to other errors in the scores.

[14]Toops, H. A. and Symonds, P. M. "What Shall We Expect of the A.Q.?" *Journal of Educational Psychology,* 13:513-28, December, 1922; 14:27-38, January, 1923.

[15]Symonds, P. M. *Measurement in Secondary Education.* New York: The Macmillan Company, 1927, p. 323-30.

[16]Foran, T. G. "The Meaning and Limitations of Scores, Norms, and Standards in Educational Measurement," *Catholic University of America Educational Research Bulletin,* Vol. 3, No. 2. Washington: Catholic Education Press, February, 1928, p. 16-19, 23-26.

[17]Of some 30,000 achievement quotients, based on a number of different achievement and intelligence tests, compiled by the writer, only about one-half of one per cent exceeded 200, and 5 per cent were above 150.

that many pupils, parents, and others falsely assume that an A.Q. of 100 is satisfactory. Because of long familiarity with the percentile marking system, in which, of course, a mark of 100 is perfect, they tend to transfer this meaning to other measures. There is no doubt that some such transfer does take place. This fact, however, does not seem to be at all a valid argument for dropping or changing the ordinary achievement quotient. The answer to it is rather that all those concerned should be educated to understand that an A.Q. of 100 represents only average or mediocre performance, and that it is not very difficult to bring about this understanding if well-planned effort to do so is put forth.

Quotient measures versus difference measures. The general judgment of those connected with the educational measurement movement as to the relative merit of differences and quotients is shown by the fact that Pintner's, the best-known and most strongly advocated, difference method[18] is rarely employed, whereas quotients are receiving widespread use. The chief reason appears to be the comparative ease of understanding achievement and mental ages and the quotients computed therefrom as contrasted with educational and mental indices and their differences. It requires considerable familiarity with test scores for an uninitiated person to form a concrete idea of what an educational index of 68, for example, or a mental index of 53, means, but it is relatively easy to understand the meaning of a mental age of twelve years and six months or of an achievement age of ten years and eight months. On the other hand, there appear to be no convincing reasons why differences are preferable to quotients.

In order to see how Pintner's differences compare with achievement quotients, Table I is given. It presents the chronological ages and educational and mental point scores of ten pupils and the ages, indices, differences, and quotients computed therefrom. The point scores given in the third and fourth columns have been transmuted into the achievement ages and mental ages given in the next two columns, and likewise into the educational and mental indices in columns seven and eight. Finally, the last two columns give the differences by Pintner's method and the achievement quotients. It will be seen that on the whole there is a fair amount of agreement between the results obtained by the two methods. Pupils A, B, C, and D, in whose cases the differences are only one, either plus or minus, have A.Q.'s in no case more than four points above or below 100. Similarly those for whom the differences are greater have A.Q.'s that differ from 100

[18]*See* p. 15.

Table I. Comparison of Differences Found by Pintner's Method with Achievement Quotients

Pupil	C.A. in Months	Educ. Point Score	Ment. Point Score	A.A. in Months	M.A. in Months	Educ. Index	Ment. Index	Diff.	A.Q.
A.........	184	70	392	186	178	47	48	−1	104
B.........	152	51	308	139	135	43	42	+1	103
C.........	144	59	361	149	152	52	53	−1	98
D.........	138	64	377	154	160	60	59	+1	96
E.........	136	42	346	129	145	46	55	−9	89
F.........	132	31	338	117	142	41	56	−15	82
G.........	128	35	225	122	112	46	43	+3	109
H.........	124	24	182	108	101	41	39	+2	107
I.........	114	18	113	102	86	42	35	+7	119
J.........	108	4	18	85	66	27	16	+11	129
Mean.....	136	40	266	129	128	45	45	0	104

by larger amounts. The relationship is by no means perfect, however. Thus pupils A and C both have differences of —1, whereas one has an I.Q. of 104, the other of 98. Also the achievement quotient of pupil F, whose difference is — 15, falls only 18 points below average, whereas pupil J, with a difference of only + 11, has a quotient 29 points above average. The coefficient of correlation between the two is .94, which means that predictions of A.Q.'s from differences, or vice versa, would be about one-third pure guesses.[19] In other words, the conclusions drawn would sometimes differ markedly according to which measure was used.

In the foregoing discussion of the relative merits of difference and quotient measures no attention has been paid to reliability. There appears to be practically no difference in the reliability of measures of the two kinds computed from the same original data; therefore, neither is to be preferred to the other on this basis. Evidence to support this statement may be found in Appendix C.

Torgerson's efficiency quotient. The efficiency quotient suggested by Torgerson[20] has been rarely, if ever, employed by others, and does not seem to deserve a better fate. The only apparent argument in its favor as compared with the achievement quotient is that one can avoid computing achievement or educational ages by using Torgerson's method. Since for many tests, all that is required to secure age norms is merely to read them off from a transmutation table, the amount of labor saved in such cases is negligible. Moreover, it is

[19]The coefficient of alienation, which equals $\sqrt{1 - r^2}$, is a measure of departure from perfect correlation. It indicates to what extent the errors involved in estimating measures or scores in one series from those in the other approach those in pure guesses. Thus in the case given above, the errors are .34 as large as they would be if the estimated scores were pure guesses; that is, if no correlation at all existed. For a more complete explanation, *see*:
Odell, C. W. "The Interpretation of the Probable Error and the Coefficient of Correlation," *University of Illinois Bulletin*, Vol. 23, No. 52, Bureau of Educational Research Bulletin No. 32. Urbana: University of Illinois, 1926, p. 41-45.

[20]*See* p. 15f.

TABLE II. COMPARISON OF TORGERSON'S EFFICIENCY QUOTIENTS WITH ACHIEVEMENT QUOTIENTS

Pupil	C.A. in Months	M.A. in Months	I.Q.	Point Score	A.A. in Months	Torgerson's A.Q.	Torgerson's Eff.Q.	A.Q.
A..........	195	190	97	70	183	108	111	96
B..........	188	163	87	102	231	157	180	142
C..........	188	140	74	54	159	85	115	114
D..........	178	162	91	78	195	120	132	120
E..........	168	164	98	69	182	106	108	111
F..........	165	151	92	34	129	52	57	85
G..........	164	178	109	73	188	112	103	106
H..........	164	192	117	31	125	48	41	65
I..........	158	164	104	78	195	120	115	119
J..........	155	205	132	113	248	174	132	121
Mean......	172	171	100	70	184	108	109	108

very often desirable to turn point scores into age scores irrespective of whether or not quotients are later to be computed. Probably the chief adverse argument is the same as that stated in the case of Pintner's method, that it is not so readily understood as the ordinary age and quotient procedure.

The differences between efficiency quotients computed by Torgerson's method and achievement quotients found by Monroe's or the ordinary method may be shown by the figures for ten eighth-grade pupils given in Table II. The first column of figures gives the chronological ages of the pupils in months. Next are their mental ages determined from intelligence test scores, and their intelligence quotients computed, of course, by dividing the mental ages by the chronological ages. In the next column are their point scores upon a subject-matter test and then the achievement ages equivalent to these point scores. Following these are what Torgerson calls achievement quotients, found by dividing each pupil's point score by the norm for the grade, which in this case is 65. The next to the last column contains Torgerson's efficiency quotients obtained by dividing his achievement quotients by the intelligence quotients. Finally come achievement quotients computed in the usual manner, that is, by dividing achievement ages by mental ages.

It will be seen that for this small group of pupils with a mean I.Q. of 100, the two methods yield approximately the same average results, the average Eff.Q. being 109 and the average A.Q. 108. A comparison of the quotients of the individual pupils, however, reveals some rather large differences, even though the coefficient of correlation between the two is .97. For pupils C, E, G, and I the differences are less than five points, but for others they are much larger, those for B, F, and H being 24 or more points. The chance element[21] in

[21] *See* footnote on p. 27.

predicting one from the other is almost one-fourth. Furthermore, it appears that on the whole this method has the undesirable effect of tending to increase high quotients and decrease low ones. Although the extreme A.Q.'s are only 65 and 142, the extreme efficiency quotients are 41 and 180. Also pupil F, with an A.Q. of 85 has an Eff.Q. of only 57, whereas pupil D's 120 and pupil J's 121 both become 132. It happens that in the case of the achievement test of which the scores are used in this illustration, the relationship between point scores and achievement ages is rectilinear. In other words, a certain difference in point score is always equal to the same difference in achievement age regardless of where it occurs, one point equalling one and one-half months. In the case of an achievement test concerning which this is not true, but for which the line of relationship is curvilinear, the differences between ordinary achievement quotients and Torgerson's efficiency quotients would tend to be even greater and more irregular than those in the example above.

Peters' accomplishment quotient. Although it is true, as Peters points out, that the ordinary accomplishment quotient is not very satisfactory for high-school use because the age norms upon which it must be based are unsatisfactory, yet Peters' proposal[22] does not seem to meet the need. The chief objection to it is that it is too complex and difficult for common use. Research workers and others engaged in experimentation might employ it, but the ordinary classroom teacher, supervisor, or administrator can hardly be expected to do so.

A second, though less important, objection is that the formula given by Peters is inapplicable to certain cases, and in others yields results which are evidently not valid. The reason this objection is not more important is that such cases are decidedly unusual. However, there would probably be a few of them in every group of several hundred or more students, and perhaps one in almost every class. It will be recalled that the formula which Peters recommends is

$$\text{A.Q.} = \frac{A}{I} \pm \frac{.13}{(5 \pm 4 - I)\ (5 \pm 4 - A)},$$

in which A equals the academic mark in standard deviation units, I the intelligence test score in such units, the plus sign is to be used if I is greater than five, and the minus sign if it is less than five. If either A or I in the formula becomes nine or one, the denominator in the fraction becomes zero and, therefore, the value of the fraction and also of the A.Q. equals infinity. In case both A and I are slightly

[22] *See* p. 16f.

above or below 1.00, the result is a negative achievement quotient, which is patently impossible. For example, if A equals .70, and I equals .80, the A.Q. value given by the formula is — 1.29.[23] If both A and I have values slightly above or slightly below 9.00, the formula yields a ridiculously large result. For example, if A equals 9.2, and I equals 9.1, the resulting A.Q. is 7.51, which is manifestly absurd, especially in view of the fact that the difference between A and I is only .1.

Otis' accomplishment quotient. The method suggested by Otis and described by Peters[24] is open to the same chief objection as is that proposed by Peters himself. In other words, it is too difficult for regular classroom use. Moreover, from the statistical standpoint it appears to be inferior to Peters' proposal. After presenting it, Peters goes on to say that although it has statistical plausibility, there are several serious objections. First, accomplishment quotients according to it would have no standard meaning, since their range depended upon the degree of correlation between intelligence and achievement. If the test were perfect, the student at the top in intelligence could not secure an A.Q. of more than 1.00. Second, if the mean accomplishment corresponding to a particular degree of intelligence is taken as normal for that degree, the assumption is made that the whole lag is due to lack of effort, but as a matter of fact most of it is due to the inferiority of the test. We should divide a measure of what one does achieve by a true measure of what he is able to achieve, which is not done by Otis' formula. Third, gratuitous additions to A.Q.'s are made for no satisfactory reason. For example, if $r = .60$, a student whose intelligence and achievement were both $+1\sigma$ would get an A.Q. of 1.07; if they were both 2σ, of 1.11; whereas in both cases he should, of course, have an A.Q. of 1.00. Fourth, the method is clumsy, because a new formula would have to be made for every different test used and every different school that employed the method. Furthermore, the empirical adjustment to secure zero correlation between A.Q. and I.Q. renders the procedure too difficult to expect its common use. Fifth, Peters' method is closely analogous to the usual one for A.Q.'s, whereas that of Otis gives a new meaning to the resulting quotient.

It seems to the writer that on the whole Peters' objections are well founded. Probably the most weighty of these objections is that the adjustment to secure the desired zero correlation is decidedly

---

[23]It will be recalled that Peters does not omit the decimal point in writing the accomplishment quotient, so that —1.29, for instance, is the same as the more usual —129.

[24]*See* p. 17f.

difficult, and also that a new formula must be made for each different test and school. Very few persons in ordinary public-school work would be willing to go to this amount of trouble to compute quotients.

Symonds' index of effort. The index of effort or of studiousness suggested by Symonds[25] is a rather general expression that may be applied to any measure which accomplishes the desired result. Of the two methods he suggests for computing it, the first is a fairly good one for rough work. The index found by it is easily computed and understood. On the other hand, the same objection applies to it as to all other measures based upon mere ranks rather than upon exact scores. This is that if two pupils rank next to each other, it makes no distinction according to the size of the difference. For example, if the best two pupils in one group have scores of 48 and 47, respectively, and the best two in another group of 48 and 40, the differences in rank will be the same although the difference in scores is eight times as great in the second case as in the first. Also, indices secured by this method do not mean the same when different numbers of pupils are concerned. An index or difference of a given amount is much more likely to occur if the pupil for whom it is computed is one of a large group than if he belongs to a small group. That is to say, the difference between first and second rank in a group of ten is ordinarily much greater than in a group of 50 or 100. On the average, the difference between the same two ranks in groups of different size varies inversely as the size of the groups. Hence, in general, the difference between ranks 1 and 2 in a group of ten would be five times as great as in a group of fifty. The writer recommends, therefore, that if this index is used, the difference in ranks be divided by the number of cases and the result multiplied by 100, to eliminate decimals. In formula form, $\text{Index} = \frac{R_A - R_I}{N} 100$, in which R_A equals rank[26] in achievement, R_I in intelligence, and N the number of individuals in the group.

The second method suggested by Symonds is considerably more difficult than the first and not readily understood by persons who have not had some statistical training. It is, however, not as complicated as one or two other methods that have been suggested, and from a statistical standpoint appears to be sound. However, the writer doubts if the rank and file of teachers or even of supervisors and administra-

---

[25] *See* p. 18f.

[26] In using this formula, a large value of R indicates high rank, and a small value low rank. For example, in a group of 25 pupils, rank 25 denotes the best, rank 24 the second best, and so on down to 1, which denotes the worst.

TABLE III. COMPARISON OF INDICES OF EFFORT COMPUTED BY THE TWO METHODS SUGGESTED BY SYMONDS

Pupil	Ach. Point Score	Intel. Point Score	Ach. Rank	Int. Rank	Ach. P.S. Dev.	Intel. P.S. Dev.	Ach. S.D. Dev.	Intel. S.D. Dev.	Diff.	S.D. Index	Diff. Index
A.....	36	178	10	8	+8	+19	+1.5	+.9	+.6	56	+2
B.....	33	166	9	6	+5	+7	+1.0	+.3	+.7	57	+3
C.....	32	181	8	9	+4	+22	+.8	+1.0	−.2	48	−1
D.....	30	165	7	5	+2	+6	+.4	+.3	+.1	51	+2
E.....	29	186	6	10	+1	+27	+.2	+1.2	−1.0	40	−4
F.....	28	170	5	7	0	+11	0	+.5	−.5	45	−2
G.....	27	154	4	4	−1	−5	−.2	−.2	0	50	0
H.....	26	138	3	2	−2	−21	−.4	−.9	+.5	55	+1
I.....	22	142	2	3	−6	−17	−1.2	−.8	−.4	46	−1
J.....	17	110	1	1	−11	−49	−2.1	−2.2	+.1	51	0
Mean.	28	159	5.5	5.5	0	0	0	0	0	50	0

tors can be brought to use a method that requires the computation of the standard deviation and the transmutation of scores or differences into standard deviation units.

In order that these two indices may be compared, Table III is given. Near the left of this table may be found the achievement and intelligence point scores of ten pupils. The next two columns contain their ranks, 10 being highest and 1 lowest. The next pair contain the deviations from the means, which are, respectively, 28 and 159. These deviations have been divided by the standard deviations, 5.2 for achievement and 22.2 for intelligence, and the results entered in the next two columns. Under the heading "Diff." may be found the differences between the entries in the two preceding columns. These differences are then multiplied by ten and added algebraically to fifty, as called for by Symonds' second method, and the results given in the next to the last column. The last column contains the differences according to his first method, these being found by subtracting the intelligence rank from the achievement rank of each pupil. A comparison of the last two columns shows a tendency toward agreement, although it is not perfect. For example, pupils D and J both have indices of 51 according to the second method, but the first has + 2 and the second 0 by the other. Likewise, pupils A and D both have indices of + 2 by the first, but 56 and 51, respectively, by the second. The coefficient of correlation in this case is .95; hence the chance element[27] in predicting either index from the other is about one-third. No attempt has been made to compare Symonds' suggested measures with achievement quotients, since the latter are commonly not found for high-school subjects. Indeed, Symonds does not recommend his in-

[27]*See* footnote on p. 27.

dex as better than the quotient, but only as usable where the latter is not.

Nygaard's accomplishment quotient. In support of his proposal that $A.Q. = \frac{A.A.}{\text{Predicted } A.A.}$, instead of $\frac{A.A.}{M.A.}$,[28] Nygaard advanced several arguments. One of these is that the average A.Q. of any group will be 100 irrespective of how it ranks. This is true, but not, therefore, necessarily desirable. A.Q.'s computed on this basis would allow valid comparisons to be made within the class or other group in question, but not between members of it and those of other groups. For some purposes, it may be desirable to determine how well pupils are doing with regard to all factors that condition learning, such as effects of home training, subject-matter studied, teacher's ability, and habits of study. Such a measure as Nygaard's A.Q. would determine this rather well, but it would not at all show how well a teacher was realizing on the capacities of her pupils. If because of very poor instruction, or, for that matter, of any other reason, her class as a whole was doing very poor work, A.Q.'s computed according to Nygaard's method would not reveal this fact. On the whole, it seems much more desirable to make use of A.Q.'s that permit valid comparisons between all pupils regardless of whether they are in the same small group or not.

A second argument advanced by Nygaard is that the negative correlation between A.Q.'s and I.Q.'s will be eliminated.[29] To the writer, this also seems to possess little or no validity. It is, indeed, generally recognized that it would be desirable to alter the fact that on the whole bright pupils do poorer work in relation to their capacity than do dull pupils, but the mere use of a statistical method or device which eliminates this negative correlation without changing the actual quality of the work done appears to have no merit. Indeed, it may be argued that it is positively undesirable, since it tends to conceal true conditions. Rather than to attempt to alter the correlation between the A.Q. and the I.Q. by statistical devices, one should follow the suggestion made by Torgerson,[30] as well as others. This is that the negative correlation be eliminated by so classifying or grouping

---

[28]*See* p. 18f.

[29]In order to compare the correlation between Nygaard's A.Q. and the I.Q. with that between the ordinary A.Q. and the I.Q., the writer had both computed for a group of several hundred fourth-grade pupils. The coefficient of correlation between ordinary A.Q.'s on the Stanford Arithmetic Test, Form A, and I.Q.'s computed from the National Intelligence Tests, Scale A, Form 1, was —.42 and that between Form B and Form 2, respectively, was —.58. These coefficients indicate very well the general trend of the many reported correlations between ordinary A.Q.'s and I.Q.'s. The corresponding correlations for Nygaard's A.Q.'s with I.Q.'s were in both cases very small and positive, being respectively .12 and .07.

[30]Torgerson, T. L. "Is Classification by Mental Ages and Intelligence Quotients Worth While?" *Journal of Educational Research,* 13:171-80, March, 1926.

pupils, so planning the work they are to carry, and so instructing them, that all come as near as possible to realizing their capacities to the fullest extent, and that in so far as this ideal goal is not reached, it be approached to approximately the same degree for pupils of all degrees of intelligence.

It is a merit of Nygaard's formula that it makes allowance for the fact that the range of achievement ages of a group is usually not equal to that of its mental ages. As Miss Rand[31] and Kelley[32] have pointed out, this condition constitutes a more or less serious statistical objection to the A.Q. as usually computed.[33] It does not seem to the writer, however, that this merit of Nygaard's proposal is of sufficient importance to justify its substitution for the ordinary method.

Summary. In this chapter, the writer has dealt with certain controversial questions having to do with measures of achievement relative to capacity. In the first place, he recommends that "achievement quotient" be used for the comparison of achievement with capacity—that is, potential mental ability—and that "subject quotient" and "educational quotient" be used for the comparison of achievement with chronological age. The greater usefulness and significance of comparing achievement with mental age than with chronological age is emphasized, and certain arguments that have been advanced in favor of the latter comparison answered. It is shown that Franzen's concept of an A.R. or an A.Q. of 100 as maximum is erroneous, but that instead 100 is the average. The recommendation is made that quotient measures rather than difference measures be employed chiefly because their use is already much more common. Following this, the suggestions of Torgerson, Peters, Otis, Symonds and Nygaard are considered critically. Of these only that of Symonds, who proposed an index of effort for use above the elementary school, is considered to have much practical merit.

---

[31]Rand, Gertrude. "A Discussion of the Quotient Method of Specifying Test Results," *Journal of Educational Psychology*, 16:599-618, December, 1925.

[32]Kelley, T. L. *Statistical Method*. New York: The Macmillan Company, 1923, p. 109-14.

[33]This statistical consideration is discussed more fully on p. 36f.

CHAPTER IV

THE VALIDITY OF MEASURES OF ACHIEVEMENT RELATIVE TO CAPACITY

Problem of this chapter. Inasmuch as the validity[1] of most of the various proposed measures of achievement relative to capacity depends upon the same conditions and assumptions, it has seemed desirable to treat this question in a separate chapter and not in connection with the comparative merits of the various measures. In general, it cannot be said that any one of the measures named and discussed in the first three chapters is to be preferred to the others because it is more valid. Practically all of those suggested are subject to limitations of this sort, and it is the purpose of this chapter to point out what these are and also to suggest how they may, to some extent at least, be avoided. In other words, methods of computing such measures that avoid, or partially avoid, these limitations, will be described and criticized.

In order that measures of achievement relative to capacity be valid, it is not sufficient that the separate measures of achievement and of capacity upon which they are based be valid. Several other conditions must be met. One is that, as Sherrod[2] and Popenoe[3] have pointed out, the age norms upon which achievement quotients are based must be perfect if the quotients are to be entirely valid. In other words, the basis of transmutation of point scores into mental and achievement ages must be perfect. A further condition, pointed out by Rand[4] and Kelley[5] among others, is that the units employed in both numerator and denominator or, in other words, in both achievement age and mental age, must be equivalent. Still another objection to the validity of the A.Q. has been advanced by Goodenough.[6] She points out that the A.Q. is based upon the assumption that achievement age develops parallel with mental age from birth, whereas in reality it ordinarily lags a great deal behind mental age until the be-

---

[1]For a definition of validity *see* p. 11.

[2]Sherrod, C. C. "The Development of the Idea of Quotients in Education," *Peabody Journal of Education,* 1:44-49, July, 1923.

[3]Popenoe, Herbert. "A Report of Certain Significant Deficiencies of the Accomplishment Quotient," *Journal of Educational Research,* 16:40-47, June, 1927.

[4]Rand, Gertrude. "A Discussion of the Quotient Method of Specifying Test Results," *Journal of Educational Psychology,* 16:599-618, December, 1925.

[5]Kelley, T. L. *Statistical Method.* New York: The Macmillan Company, 1923, p. 109-14.

[6]Goodenough, F. L. "Efficiency in Learning and the Accomplishment Ratio," *Journal of Educational Research,* 12:297-300, November, 1925.

ginning of attendance at school. Wilson[7] has likewise advanced a very similar objection. Since the first of these conditions, that of perfect age norms, is a matter of achievement and mental ages rather than of quotients, it will not be further discussed here. The others, however, will be elaborated in the following paragraphs.

Necessity of equivalent units for achievement age and mental age. Probably the best general discussion of this point is by Kelley,[8] who points out that three conditions must be met before two scales are entirely equivalent. As applied to achievement quotients or similar measures, this means that achievement scores and intelligence scores must both meet these three conditions in order to make dividing the former by the latter justifiable. The three conditions are that one point of the first scale must be known to be equal to a point of the second, also a second point of the first equal to a second point of the second, and that the law of relationship between successive points must be the same for the two scales. Kelley does not make any particular application to achievement and intelligence tests, but gives other illustrations showing very clearly the hazards involved and the misleading conclusions that may be drawn if these conditions are not met.

Miss Rand's discussion[9] of this point is perhaps more concrete than Kelley's, upon whose treatment she bases hers. She stated that "We are early taught that we must not divide months by years, grams by ounces, centimeters by inches. Why, then should we divide E.Q.'s by I.Q.'s or E.A.'s by M.A.'s without proof of their equivalence . . . .?" Following this, some evidence is offered that the E.Q. unit is smaller than the I.Q. unit. This evidence includes two quotations from Burt,[10] stating her contention as a fact, and data from four or five studies. From Ruch[11] she quotes standard deviations of the I.Q. of 14.2 and similar deviations of the E.Q. of 10.4, 12.0, and 16.3. In connection with the New Jersey Composite Test,[12] she cites figures indicating that the educational test unit is only about two-thirds that of the intelligence test unit. For the Lippincott-Chapman Classroom Products Survey Tests[13] she computed standard deviations and com-

---

[7]Wilson, W. R. "The Misleading Accomplishment Quotient," *Journal of Educational Research*, 17:1-10, January, 1928.
[8]Kelley, *loc. cit.*
[9]Rand, Gertrude, *loc. cit.*
[10]Burt, Cyril. *Mental and Scholastic Tests.* London: P. S. King and Company, 1922, p. 158 and 176.
[11]Ruch, G. M. "The Achievement Quotient Technique," *Journal of Educational Psychology*, 14:334-43, September, 1923.
[12]"Preliminary Standardization of the New Jersey Composite Test." Trenton, New Jersey: Department of Public Instruction, 1923. 13 p.
[13]Chapman, J. C. "Lippincott-Chapman Classroom Products Survey Tests." Philadelphia: J. B. Lippincott Company, 1920.

pared them with those for the Terman Group Test of Mental Ability[14] showing that the ratio between the two is approximately nine to thirteen. Finally, a reference is given to Kelley,[15] who reported a smaller standard deviation for E.Q.'s from the Stanford Achievement Test than is commonly found for I.Q.'s in the same grade. Miss Rand's conclusion appears to be that accomplishment quotients should not be employed, because their erroneous interpretation and use more than outweighs their practical value.

Miss Rand's contention that the units employed in expressing achievement and intelligence test scores are not equivalent is, in the opinion of the writer, entirely justified. Nygaard[16] recognized it when he urged as one of the merits of his proposal that it would correct for this lack of equivalence. Certain data which the writer has compiled tend not only to confirm Miss Rand's position, but to make the situation appear even worse than she has portrayed it. The standard deviations of achievement ages computed from scores upon the Stanford Arithmetic Test, the Monroe General Survey Scale in Arithmetic, and the Monroe Standardized Silent Reading Tests, Revised, differ considerably from those for the National Intelligence Tests, Scale A, and the Illinois General Intelligence Scale administered to the same pupils. The ratio of the standard deviation of the Stanford Arithmetic Test to that of the National Intelligence Scale was found to be about three to four. For the other subject-matter tests mentioned, the standard deviations were approximately twice as great as that for the Illinois General Intelligence Scale. Since these results are based upon two testings of several hundred pupils in the case of each test, they may be considered fairly reliable. Certainly the differences are great enough that the statement seems justified that the units do differ so greatly as not to approach equivalence, although all the differences are not in the same direction. Moreover, the fact that most of the differences are in the opposite direction from those reported by Miss Rand, as well as the fact that all those reported by the author do not agree, makes the situation even worse with regard to the A.Q., since it appears that in some cases the unit in the numerator is the greater, in others that in the denominator is the greater.

To remedy the situation, Miss Rand proposes a program of reconstruction including two plans. The first is that all test scores be expressed in terms of "an arbitrary scale of values having a fixed

[14]"Data on the Repetition of Certain Mental Tests," *Journal of Educational Research*, 7:458-59, May, 1923.

[15]Kelley, T. L. "A New Method for Determining the Significance of Differences in Intelligence and Achievement Scores," *Journal of Educational Psychology*, 14:326, September, 1923.

[16]*See* p. 19.

zero and a fixed scale number or unit which is to be applied to each σ or fraction of σ of score above the zero." She refers with apparent approval to a suggestion of Rugg's[17] that such a scale range from -2.5σ as zero, to $+2.5\sigma$ as 100, thus letting ten scale points represent each $.5\sigma$. She mentions the possibility of employing McCall's T-scale,[18] but states her opinion that it would be better to make a suitable scale for each age at which a test is being used rather than to employ a scale based on twelve-year-old performance for all ages.

Her second suggestion is that quotients for other tests should be arbitrarily made to conform to Stanford-Binet I.Q.'s by the proper statistical transmutations to make them equivalent thereto. These, of course, would be based on the ratios of the Stanford-Binet I.Q.'s to those of the tests in question. For this procedure also, she suggests that the proper basis of transmutation for each age be determined and used.

Miss Rand's proposals appear to be statistically satisfactory, but it is doubtful if they will ever receive wide use because of the amount of computation necessary. Any measure that is to receive general acceptance must be fairly simple to compute and understand, and hers suffer decidedly in these respects by comparison with the simple achievement quotient and other measures.

Goodenough's and Wilson's argument against the validity of the A.Q. As has been mentioned already,[19] Miss Goodenough has shown that the achievement quotient as ordinarily employed is not entirely valid. She centered her attack on the point that "the accomplishment ratio does not afford a valid means for comparing the learning efficiency of individuals or of groups who differ widely in intelligence." Table IV, taken from her discussion, shows the various rates at which bright pupils must do their work to earn accomplishment ratios of 100 at various mental ages. From this it can easily be seen that in order to maintain equal A.R.'s, bright pupils must learn at a much greater rate than dull pupils. For example, a pupil with a mental age of 10 and an I.Q. of 120 has, on the average, been in school only 4 semesters, whereas one of the same mental age but with an I.Q. of 80 has attended 12 semesters or three times as long.[20] Therefore, if the two started at the same point when they entered school, the bright-

[17]Rugg, H. O. *Statistical Methods Applied to Education.* Boston: Houghton Mifflin Company, 1917, p. 222 and 392.

[18]*See* p. 8.

[19]*See* p. 35f.

[20]The first pupil mentioned has a chronological age of $\frac{10}{1.20}$ or 8⅓ years, the second of $\frac{10}{.80}$ or 12.5 years. Thus, the former has probably spent 2 years, or 4 semesters in school, the latter 6 years, or 12 semesters.

Table IV. Length of School Attendance at Various Mental Ages for Children of Different Intellectual Levels[a]

Intelligence Quotient	Length of Attendance in Semesters if Mental Ages are as Indicated				
	8–0	10–0	12–0	14–0	16–0
140	0	2	5	7	10
120	1	4	7	11	14
100	3	7	11	15	19
80	7	12	17		
60	14				

Table reads: A child whose intelligence quotient is 140, if he has a mental age of 8, will normally have had no school attendance; if he has a mental age of 10, he will have attended 2 semesters; if he has a mental age of 12, he will have attended 5 semesters; etc.

[a]Goodenough, F. L. "Efficiency in Learning and the Accomplishment Ratio," *Journal of Educational Research*, 12:299, November, 1925.

er one must cover the work three times as fast as the duller one, although his I.Q. is only one and one-half times as great.

Wilson[21] has given attention to this same point, that the achievement quotient is not valid for comparing pupils of different degrees of ability. In his discussion he offers an elementary proof that the accomplishment ratio method, as he calls it, results in a spurious increase of the A.Q.'s of pupils of below average ability, and a decrease for those above. To prove this, Wilson took the 48 pupils in Franzen's group and assumed that their obtained I.Q.'s were perfectly accurate and that their true efficiency quotients were the same as their intelligence quotients; in other words, that the true accomplishment quotient of each pupil was 100. On the assumption of a probable error of three points, which is much smaller than is usually found in actual practice, for both I.Q. and E.Q., and a random distribution of errors, he obtained intelligence and efficiency quotients such as might be secured in actual testing and from them computed accomplishment quotients. Although the true A.Q.'s were all 100, those obtained by his method ranged from 85 to 117. Furthermore, although the correlation between true A.Q.'s and I.Q.'s was zero, that for the obtained A.Q.'s and I.Q.'s was —.38. Also he gave a simple geometrical proof that the correlation between actually obtained I.Q.'s and A.Q.'s must be negative. He concluded that our present measures of ability and achievement are so lacking in reliability, and perhaps also in validity, that "they cannot safely be assumed to tell with definiteness anything concerning the true accomplishment quotients of the students measured."

Summary. In this chapter, several conditions prerequisite to the validity of measures of achievement relative to capacity have been

[21]Wilson, W. R. "The Misleading Accomplishment Quotient," *Journal of Educational Research,* 17:1-10, January, 1928.

stated, and it has been shown that ordinarily these are not satisfactorily fulfilled. The age norms upon which achievement quotients are based are not perfect. The units in which mental ages and achievement ages are expressed are frequently not equivalent. The implied assumption that the achievement age should develop parallel with mental age from birth is not true to the facts. One or two proposals for improving the situation by rendering achievement quotients statistically valid in so far as certain points are concerned have been given. It does not appear, however, that these proposals are likely to receive wide use because of their lack of simplicity, although either one of Miss Rand's suggested procedures would remove the chief statistical hindrance to the validity of the A.Q. On the whole, the conclusion appears inevitable that if the achievement quotient or any similar measure is to be employed, it must be only in a very cautious and general manner, since its validity is too low to justify more exact use.

CHAPTER V

THE RELIABILITY OF MEASURES OF ACHIEVEMENT RELATIVE TO CAPACITY

Problem of this chapter. It is the purpose of this chapter to present evidence and draw conclusions as to the reliability[1] of measures of achievement relative to capacity, especially of the achievement quotient. Results given by several other writers will be briefly presented, as also will be some hitherto unpublished data gathered by the present writer, and finally the conclusions that seem warranted will be stated.

The bases of reliability. Undoubtedly, the primary basis of the reliability of achievement quotients or other measures of achievement relative to capacity is that the achievement and mental scores or, in other words, the achievement and general intelligence tests upon which they are based, be themselves reliable. This has been pointed out by a number of writers, including Toops and Symonds,[2] Chapman,[3] Beeson and Tope,[4] Foran[5] and Herring.[6] It is evident, without going into the matter from a statistical standpoint, that a quotient or other quantity computed from two unreliable quantities will tend to be more unreliable than either one of them, since in many cases positive errors will be added to positive ones, and in many others, negative to negative. Such statements as Herring's[7] that "accomplishment differences are comparatively reliable when the tests employed are comparatively reliable" are not justified unless the word "comparatively" is interpreted more loosely in one place than in the other. Neither

[1]As was stated in the explanation of reliability given on p. 11, a test or measure is said to be reliable not only if a second application yields the same scores as the first, but also if there is a constant and known difference between the two sets of scores. In other words, a test is reliable if variable errors—that is, errors which are more or less accidental and differ for the different individuals concerned—are eliminated. There may be constant errors, errors that tend to be the same for the whole group, present and yet reliability be perfect or nearly so. Such causes as too long or too short time limits, practice effect from having had a similar test very recently, and so forth, tend to raise or lower, as the case may be, the scores made by all members of the group tested. This effect, of course, carries over to achievement quotients or any other measures of achievement relative to capacity, and makes them less accurate though not less reliable in the technical sense of the term. It should not be overlooked, therefore, that relative measures may contain constant errors whether or not their reliability is high. If it is very high, such errors will be practically the only ones present; if not, they will be present in addition to the variable errors.

[2]Toops, H. A. and Symonds, P. M. "What Shall We Expect of the A.Q.?" *Journal of Educational Psychology*, 13:513-28, December, 1922; 14:27-38, January, 1923.

[3]Chapman, J. C. "The Unreliability of the Difference Between Intelligence and Educational Ratings," *Journal of Educational Psychology*, 14:103-8, February, 1923.

[4]Beeson, M. F. and Tope, R. E. "The Educational and Accomplishment Quotients as an Aid in the Classification of Pupils," *Journal of Educational Research*, 9:281-92, April, 1924.

[5]Foran, T. G. "The Meaning and Limitations of Scores, Norms, and Standards in Educational Measurement," *Catholic University of America Educational Research Bulletin*, Vol. 3, No. 2. Washington: Catholic Education Press, February, 1928, p. 16-19, 23-26.

[6]Herring, J. P. "The Reliability of Accomplishment Differences," *Journal of Educational Psychology*, 15:530-38, November, 1924.

[7]*Ibid.*

can Wilson's assumption[8] that a coefficient of reliability[9] of test scores of .90 is fairly satisfactory be considered valid. For such a coefficient, the element of uncertainty or guessing[10] in the scores is about four-ninths. Since the element of error in a quotient is greater than that in either the numerator or denominator, and therefore greater than four-ninths for a reliability of .90, it can scarcely be said that such a coefficient produces a reliability of achievement quotients that is at all satisfactory. Herring[11] has suggested that the requisite reliability of test scores to yield satisfactory A.Q.'s should be .95. This is probably as high as can be hoped for with even our best tests,[12] and yet the element of guessing connected with a coefficient of correlation of .95 is almost one-third, and consequently that with the resulting quotient still greater. Ruch[13] has suggested that apparently the chief cause of too great unreliability of achievement quotients is that the tests upon which they are based are too short from the standpoint of time spent by the pupils in taking them. He says that apparently I.Q.'s should be based upon tests which require at least thirty minutes of working time. It does not appear, however, that this proposal is satisfactory, as it is well known that the reliability of most tests that consume thirty minutes or even more does not approach even .95 very closely. Practically all of the few tests which do equal or exceed this figure are either individual intelligence tests or group tests lasting two or three hours.

It is true, as Ruch points out, that one important factor in reliability is frequently the short time limit of a test. For exercises of the same sort, reliability increases as the square root of the ratio of the times. For example, the reliability of a test four times as long as another is twice as great. In practice, one of the most convenient ways of increasing reliability through devoting more time to testing is, as Beeson and Tope[14] have pointed out, to employ average measures based upon several group tests or perhaps, in the case of intelligence, scores from individual tests.

---

[8]Wilson, W. R. "The Misleading Accomplishment Quotient," *Journal of Educational Research,* 17:1-10, January, 1928.

[9]The coefficient of reliability is the coefficient of correlation between the scores secured from two applications of the same test or duplicate forms thereof within a short period.

[10]*See* footnote on p. 27.

[11]Herring, J. P. "The Reliability of Accomplishment Differences," *Journal of Educational Psychology,* 15:530-38, November, 1924.

[12]Few group tests yield reliability coefficients as high as .95, or for that matter, even as high as .90.

[13]Ruch, G. M. "The Achievement Quotient Technique," *Journal of Educational Psychology,* 14:334-43, September, 1923.

[14]Beeson, M. F. and Tope, R. E. "The Educational and Accomplishment Quotients as an Aid in the Classification of Pupils," *Journal of Educational Research,* 9:281-92, April, 1924.

Reported data on the reliability of the A.Q. Among those who have reported such data is Symonds.[15] As a result of employing the National Intelligence Tests, the Woody-McCall Mixed Fundamentals in Arithmetic, and the Thorndike-McCall Reading Scale, he found coefficients of correlation or reliability between achievement ratios from first and second forms of from .23 to .60, with probable errors[16] of 6 or 7 points. He showed that if the tests were given twice or two forms used as one, the probable error of the A.R. became slightly less than 5 points in each case, or, in other words, probably about the same as that of an I.Q. based upon one of the best individual intelligence tests. He next proceeded to consider the reliability of school marks based upon the achievement ratio, and showed that it is possible to adopt a five-letter marking system such that the probable error of the achievement ratio varies from about one-fourth to less than one-half of a letter interval. With other five and six-letter marking systems, however, the probable error of the A.R. may be as great as two-thirds of a letter interval. After comparing these figures with previously published data concerning the reliability of scores on the Woody-McCall Arithmetic and Thorndike-McCall Reading Scales for grade placement, he concluded that "the A.R. is more accurate a unit for marking than the score is accurate for placing in the proper grade." Further, he stated that the reliability coefficients of the arithmetic A.R. (.60 and .49) were comparable with those of ordinary school marks. This, however, did not hold for the reading A.Q., for which the coefficients were only .34 and .23.

Popenoe[17] also has reported some information of this sort. For 600 pupils, he found a coefficient of reliability of the A.Q. of .28 and a probable error of about six points. He mentioned also that several minor studies in which the mental age was kept constant so that only the numerator contained errors yielded coefficients of reliability of only about .50.

There have not been a great many published reports dealing directly with the reliability of achievement quotients. The two just referred to, those of Symonds and Popenoe, are two of the best, and also are thoroughly typical. Such figures as they give seem to show

---

[15]Symonds, P. M. "The Accuracy of Certain Standard Tests for School Sectioning and Marking," *Journal of Educational Psychology*, 15:423-32, October, 1924.

[16]The probable error is greater than half of the errors concerned, and less than the other half. Thus, in the example given above, a probable error of six or seven points means that the errors in half of the scores were less than this amount, and those in the other half greater.

[17]Popenoe, Herbert. "A Report of Certain Significant Deficiencies of the Accomplishment Quotient," *Journal of Educational Research*, 16:40-47, June, 1927.

that the reliability of the A.Q. is so low that comparatively little confidence can be placed in it.

The writer's data on the reliability of the A.Q. As has already been stated, the writer wishes to supplement the rather meager published data upon the reliability of the A.Q. by presenting some which he has recently compiled. These were obtained from two sources. The National Intelligence Test, Scale A, Forms 1 and 2, and the Arithmetic Examination of the Stanford Achievement Test, Forms A and B, were given to more than 200 fourth-grade children in an Illinois city. About the same time, the Illinois General Intelligence Scale, Forms 1 and 2, was administered to approximately 800 eighth-grade pupils in another Illinois city. Half of this latter group also took Forms 1 and 2 of Test II of the Monroe Standardized Silent Reading Test, Revised, and the other half took Forms 1 and 2 of Test II of the Monroe General Survey Scale in Arithmetic. In both cities, the regular classroom teachers gave the tests, thus reproducing ordinary conditions. The test scores were turned into achievement and mental ages in the regular manner, and then the achievement quotients computed. The coefficients of reliability of A.A.'s, M.A.'s, and A.Q.'s, their probable errors of measurement and certain other measures of reliability were computed and are presented in Table V.

It will be observed that the body of this table is divided into two parts. In the first, that is, the one to the left, will be found the measures of reliability of the achievement and mental ages of the pupils and in the other the corresponding measures for the achievement quotients. The first column in each half of the table contains the coefficients of correlation, in this case coefficients of reliability, between the results from the first and second forms of each test. The next pair of columns, headed "k", contain the coefficients of alienation.[18] Following these in each case may be found the probable errors of measurement.[19] The next two columns contain the probable errors of measurement divided by the means, and finally the last two contain the probable errors of measurement divided by the standard deviations.[20]

[18] *See* footnote on p. 27.

[19] The probable error of measurement is the probable error involved in estimating true scores from actual scores. For example, a probable error of measurement of 3.9, which is given in the table as that of the Stanford Arithmetic Age, means that if the pupils' arithmetic ages made upon either form of the test be taken as their true arithmetic ages, the errors in the cases of half the pupils will be less than 3.9 months and those for the other half greater than this amount. It is ordinarily computed by the formula, $P.E._{meas.} = .6745\ \sigma\sqrt{1-r}$, in which r is the coefficient of reliability—that is, of correlation—between two forms of the same test.

[20] Since the significance of a probable error of a given amount depends to a considerable extent upon the size of the measure for which the error has been found, it is generally recommended that the probable error of measurement be divided by such a divisor that the result in any one case may be compared with that in any other. For example, a probable error of measurement of six months might be found for the achievement ages of primary

Table V. Measures of Reliability of Age and Quotient Scores Obtained by the Writer

	Age Scores					Quotient Scores				
	r	k	$P.E._{meas.}$ in months	$\frac{P.E._{meas.}}{M}$	$\frac{P.E._{meas.}}{\sigma}$	r	k	$P.E._{meas.}$ in points	$\frac{P.E._{meas.}}{M}$	$\frac{P.E._{meas.}}{\sigma}$
Stanford Arithmetic	.76	.65	3.9	.03	.33	.54	.84	4.6	.05	.45
Monroe Arithmetic	.67	.74	16.0	.07	.39	.53	.85	9.6	.08	.47
Monroe Reading Comprehension	.61	.79	15.1	.08	.42	.39	.93	9.2	.09	.53
Monroe Reading Rate	.69	.72	22.1	.10	.38	.58	.81	13.0	.11	.44
National Intelligence	.82	.57	4.7	.04	.29	...	...	...	...	...
Illinois Intelligence	.77	.64	7.1	.04	.32	...	...	...	...	...

It will be seen from the coefficients of correlation and also of alienation given in the table that the reliability of the achievement and mental ages is not very high. In the case of the achievement tests, the correlations range from .61 to .76, and the corresponding coefficients of alienation from .65 to .79. That is, the guessing element[21] ranges from about two-thirds to four-fifths. In the case of the two intelligence tests, the situation is somewhat better, although even with them the coefficients of alienation are as large as .57 and .64. The corresponding coefficients of correlation for the achievement quotients range from .39 to .58 and those of alienation from .81 to .93. In other words, in the very best case shown by these data, the guessing element involved in the achievement quotient is slightly greater than four-fifths, whereas in the worst case it is well above nine-tenths. The columns containing the probable errors of measurement and the quotients of these errors divided by the means and standard deviations, respectively, show the same tendencies in a different manner. For the Stanford Arithmetic Test they are relatively small, the probable error of the quotients obtained upon the test being less than five points, or, in other words, less than five per cent of the mean quotient and also less than half of the standard deviation of the quotients. For the Monroe Arithmetic and Reading Tests the actual probable errors of measurement are from two to three times as large as in the case of the Stanford Arithmetic, but when taken relative to the means of these tests, they are only from about one and one-half to two times as great. When divided by the standard deviations, they are nearly the same. Evidently, probable errors which amount to close to one-tenth of the mean scores are so serious that little dependence can be placed upon the reliability of the quotients. To compare the situation with a more familiar example, it is the same as if the probable error of measurement involved in measuring the heights of adults were six or seven inches. It is readily seen that for almost all purposes measurements of height of which half were in error by more than six or seven inches would be practically worthless. It seems to the writer, therefore, that from these data one must conclude that achievement quotients based upon a single administration of the tests used are so much in error that one is rarely if ever justified in employing them for individual diagnosis.

---

pupils and one of eight months for those of upper-grade pupils. However, since the actual age scores of the latter will be ordinarily about twice as great as those of the former, the relative error in their case will be smaller than in that of the primary children. The two quantities which have been suggested as divisors are the mean and the standard deviation. Each has certain advantages over the other and neither is perfect; therefore it has seemed best to employ both.

[21] *See* footnote on p. 27.

The reliability of average A.Q.'s for groups of pupils. The data so far presented, both from other sources and from the writer's investigation, show that the A.Q.'s of individual pupils are decidedly unreliable, but do not deal with average A.Q.'s of groups of pupils. The conclusion just stated as to individual A.Q.'s, however, does not warrant the same conclusion concerning average A.Q.'s of groups. As has been pointed out by Kelley,[22] Myers,[23] and others, average A.Q.'s of groups are, on the other hand, ordinarily rather reliable. This conclusion is, of course, merely the application of an elementary statistical principle or formula, that the reliability of a measure increases in direct proportion to the square root of the number of cases[24] upon which it is based. This decidedly important point has apparently been overlooked by most of those who have written concerning the reliability of the A.Q. This is unfortunate since occasions frequently arise in which it is desirable to employ the average A.Q. of a class or some other group without making any use of the A.Q.'s of its individual members.

This conclusion is supported by the writer's data shown in Table V. In the ordinary elementary school there will be very few classes as small as twenty-five and in many cases, they will be at least as large as thirty-six. Therefore, in accord with the statistical principle stated in the last paragraph, the reliability of average achievement quotients for usual elementary-school classes will be at least five or six[25] times as great as that of individual achievement quotients as shown in Table V. Dividing probable errors of measurement given in this table by five or six yields such errors of only one or two points, or, expressed otherwise, of about 1 or 2 per cent of the means, or one-tenth of the standard deviations. Errors of this magnitude are small enough that one is justified in placing considerable reliance upon the accuracy of average A.Q.'s. If differences in the average A.Q.'s of classes of approximately five points are found to exist, it will ordinarily be rather safe to assume that they indicate real differences in the degree to which the classes as wholes have capitalized their capacity to learn. For differences smaller than this, such a conclusion, if drawn at all, should ordinarily be merely tentative.

---

[22]Kelley, T. L. *Interpretation of Educational Measurements.* Yonkers: World Book Company, 1927, p. 7-8, 22-25.

[23]Myers, C. E. "The Accomplishment Ratio," *Research Bulletin of the Pennsylvania State Education Association,* No. 3. Harrisburg, Pennsylvania: Pennsylvania State Education Association, January, 1928, p. 38-40.

[24]The statement made above may perhaps be made clearer by an illustration. Since the square root of 25 is 5, the average A.Q. of a group of 25 pupils will be five times as reliable as the A.Q. of one of the individual pupils composing the group. Therefore, if the average error in the individual A.Q.'s is, let us say, ten points, that in the group A.Q. will be only one-fifth as great, or two points.

[25]Five and six are the square roots of 25 and 36, respectively.

Summary. The reliability of an achievement quotient or other measure of achievement relative to capacity is primarily determined by the reliability of the original measures upon which it is based, and is less than that of either one of the two measures unless one or both happens to possess perfect reliability, which, of course, is never true. Very few of our standardized tests possess high enough reliability that the errors involved in achievement quotients computed from scores thereon may safely be neglected. A comparatively few data previously reported indicate that the coefficient of reliability of achievement quotients is probably in most cases below .50, and its probable error at least six or seven points. Data collected by the writer likewise yield results in entire agreement with these conclusions. Thus it may be said that all the available direct data as to the reliability of the A.Q. support the inferences drawn from the low reliability of tests, that only in very exceptional cases are the A.Q.'s of individual pupils reliable enough to furnish safe guides in dealing with such pupils. On the other hand, the average A.Q.'s of classes or larger groups containing a number of pupils are probably reliable enough that we are justified in employing them as measures of the achievement of the group as a whole.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Because the use of various measures of achievement relative to capacity has been so widespread and also in general so non-critical, it has seemed worth while to prepare a critical study of such measures. A brief account of the origin of the various suggested measures of this sort is followed by discussions of the merits and demerits of the different measures and of the validity and reliability of such measures in general. The following of the proposed measures have received fairly wide use: educational quotient, subject quotient, achievement quotient or accomplishment quotient, and accomplishment ratio. Pintner's difference method, Torgerson's efficiency quotient, Peters' high-school and college accomplishment quotient and Otis' similar measure, Symonds' index of effort, Nygaard's modified accomplishment quotient and Rand's suggested program of reconstructing such measures, have received either no use at all, except perhaps by their originators, or a comparatively small amount.

The writer recommends that "quotient" be used rather than "ratio" and, therefore, "achievement quotient" or "accomplishment quotient" be applied to achievement or accomplishment age divided by mental age; that "subject quotient" and "educational quotient" be employed when the divisor is chronological age; that Symonds' "index of effort" be employed where satisfactory quotient measures are not available; and that the other suggested measures be dropped either because they require too elaborate computation or because they are not needed. The comparison of achievement with mental age is more significant than that with chronological age. Franzen's concept of an A.Q. of 100 as the theoretical maximum is erroneous; an A.Q. of this size is just average. The validity of most quotient measures is not very high, chiefly because the units in the numerator and denominator are not equivalent. The proposals for improving this condition appear to be too complicated to receive general use.

A review of all known studies of the reliability of measures of achievement relative to capacity leads to the conclusion that their reliability is decidedly unsatisfactory. This is supported by original data obtained and presented by the writer. Indeed, their reliability is so low that it is recommended that they never be employed for the diagnosis, classification, or other treatment of individual pupils except

possibly when they are based upon the combined results from several group tests or one individual intelligence test. Relative measures, for a class or larger group do, however, possess high enough reliability to warrant their use.

APPENDIX A

THE REDUCTION OF NEGATIVE CORRELATION BETWEEN INTELLIGENCE QUOTIENTS AND ACHIEVEMENT QUOTIENTS

As has been shown in the body of this bulletin[1] negative correlations are almost always found to exist between achievement quotients or similar measures and intelligence quotients. In other words, instruction is such that pupils of superior capacity are not stimulated to approach their maximum achievement as nearly as are pupils who are less highly endowed. It has been suggested, and indeed in some cases shown, that such negative correlations can be lessened, perhaps even reduced to zero; that is, to a figure which indicates that pupils of all levels of capacity are capitalizing their potential abilities to an equal degree.

Probably the first writer to call attention to this possibility was Franzen. As a part of his study of the accomplishment ratio,[2] he made an experimental attempt to motivate a rather small group of pupils so as to raise their accomplishment ratios to the maximum, which he erroneously considered to be 100.[3] Although Franzen concluded that it is possible to motivate pupils so that their A.R.'s will approach 100 rather closely, Wilson[4] has shown that this conclusion is unjustified. Using Franzen's own figures, he shows that the negative correlation between I.Q.'s and A.R.'s was not significantly changed by two years' stimulation of the pupils.

Another worker who has discussed this point is Torgerson,[5] who emphasized the point that "proper grade placement . . . . tends to raise the accomplishment quotient of all pupils to a normal maximal efficiency." In support of this he cited data secured from a study including several hundred pupils. These data showed that as the pupils were placed in their grades, those who were retarded had a median accomplishment quotient of 100; those normally placed, of 107; and those accelerated, of 118. Furthermore, they showed an average accomplishment quotient of 107 for pupils with intelligence quotients below 90, of 101 for those with I.Q.'s from 90 to 109, and of 93 for

---

[1]*See* p. 23f.
[2]*See* p. 14.
[3]The fallacy of Franzen's contention that the maximum A.R. or A.Q. is 100, is shown on p. 24f.
[4]Wilson, W. R. "The Misleading Accomplishment Quotient," *Journal of Educational Research*, 17:1-10, January, 1928.
[5]Torgerson, T. L. "Is Classification by Mental Ages and Intelligence Quotients Worth While?" *Journal of Educational Research*, 13:171-80, March, 1926.

those with I.Q.'s above 110. This evidence agrees very well with what others have found. Torgerson, however, gave further data to show that for pupils properly graded, the average A.Q.'s for the three groups according to intelligence quotients were, respectively, 108, 109, and 112, or, in other words, that the differences between them were very small. He concluded, therefore, "that when pupils are properly graded the inverse relationship between intelligence quotient and accomplishment quotient disappears."

Popenoe,[6] on the other hand, has cited some figures which do not support Torgerson's argument. Twenty-four elementary schools in which there were larger than average negative correlations between A.Q.'s and I.Q.'s were chosen and attempts made to reduce the negative correlation. Subsequent testing, however, indicated that this correlation, which averaged —.59 at the beginning of the experiment, was not materially changed. Despite Popenoe's findings, however, it appears that proper grade placement, satisfactory motivation, and teaching methods adapted to pupils' abilities will generally result in reducing the negative correlation between intelligence quotients and accomplishment quotients to practically zero.

[6]Popenoe, Herbert. "A Report of Certain Significant Deficiencies of the Accomplishment Quotient," *Journal of Educational Research,* 16:40-47, June, 1927.

APPENDIX B

ESTIMATING THE RELIABILITY OF ACHIEVEMENT QUOTIENTS FROM THAT OF ACHIEVEMENT AND MENTAL AGES

It sometimes happens that it is convenient to be able to estimate the reliability of achievement quotients without actually computing measures thereof directly from the quotients themselves. The writer, therefore, has attempted to discover a method of doing so when the reliabilities of achievement and mental ages are known. So far as he was able to learn, no valid formula for this purpose has been devised and published. There are, of course, well established formulae for measuring the reliability of index numbers and other quotients in which all of a series have the same denominator, and also for certain other expressions somewhat similar to the A.Q. A reasonably diligent and exhaustive search, however, failed to reveal any formula entirely appropriate to the purpose under discussion. The writer did, however, discover two methods, one somewhat better than the other, by which the reliability of achievement quotients may be estimated approximately when those of achievement and mental ages are known. Both of these deal with probable or standard errors. Nothing based upon coefficients of reliability was found.

One of these two methods involves the use of a standard formula for the error of a quotient.[1] This formula is as follows, expressed in symbols ordinarily used in educational work:

$$\text{P.E.}_{\cdot\frac{x}{y}} = \frac{\sqrt{\left(\dfrac{X \cdot \text{P.E.}_{\cdot y}}{Y}\right)^2 + \text{P.E.}_{\cdot x}^2}}{Y}$$

Thus if, in the case of a particular pupil, the probable error of his achievement age (X), and also that of his mental age (Y), are known, it is possible to compute that of his achievement quotient $\left(\dfrac{X}{Y}\right)$. This formula would require a separate computation and yield a different probable error for each pupil except in the case of two or more whose achievement ages and mental ages were the same. However, this is not quite what is desired in ordinary work with age and quotient

[1]The formula may be found in:
Mellor, J. W. *Higher Mathematics for Students of Chemistry and Physics.* London: Longmans, Green and Company, 1909, p. 529.

TABLE VI. COMPARISON OF PROBABLE ERRORS OF ACHIEVEMENT QUOTIENTS COMPUTED BY THE FORMULA $\text{P.E.}_{\frac{x}{y}} = \dfrac{\sqrt{\left(\dfrac{X \cdot \text{P.E.}_y}{Y}\right)^2 + \text{P.E.}_x^2}}{Y}$ WITH THOSE COMPUTED FROM THE QUOTIENTS THEMSELVES

	$\text{P.E.}_{\text{meas.}}$ in months		$\frac{\text{P.E.}_{\text{meas.}}}{M}$		$\frac{\text{P.E.}_{\text{meas.}}}{\sigma}$	
	By Formula	From Quotients	By Formula	From Quotients	By Formula	From Quotients
Stanford Arithmetic...	4.8	4.6	.05	.05	.44	.45
Monroe Arithmetic...	9.6	9.6	.08	.08	.55	.47
Monroe Reading Comprehension...	8.9	9.2	.09	.09	.54	.53
Monroe Reading Rate.	12.5	13.0	.11	.11	.54	.44

scores. The desideratum is rather a formula which may be solved just once to yield a single probable error that applies to the quotient scores of the whole group of pupils in question.

Apparently the most satisfactory way of attempting to procure a measure of the kind mentioned from this formula is to substitute in it the average achievement and mental ages of the group rather than those of an individual pupil, and thus secure a probable error applicable to the achievement quotients of the whole group. This has been done by the writer and the results given in Table VI. The first pair of columns in this table compare the probable errors of measurement by the formula with those actually computed from the quotients.[2] The second pair of columns does the same for the ratios of the probable errors of measurement to the means and the third for their ratios to the standard deviations. In applying the formula in the cases of both these ratios, it is slightly modified by dropping the denominator Y. This is necessary because the probable errors have already been expressed in terms of the means and standard deviations respectively and, therefore, should not again be divided by the mean. It will be seen that in the first pair of columns the probable errors given by the formula approach those actually computed rather closely. The largest difference, which occurs in the case of Monroe Reading Rate, is only about 4 per cent of the size of the P.E. In the case of the ratios of the probable errors to the means, the agreement is exact to two decimal places. In the case of their ratios to the standard deviations, the differences are somewhat greater, running up to almost one-fourth of

[2]The latter have already been given in Table V, but are repeated here.

the ratios themselves. It appears from these comparisons that, in the case of the data dealt with, the application of the formula given above to mean scores yields probable errors of measurement and ratios of such errors to the means accurate enough for all practical purposes. This statement can not be made, however, for the ratios of the probable errors to the standard deviations.

It is perhaps needless to say that the present study is too limited in its scope for these results to be taken as definite proof that the values yielded by the formula will always approach those actually obtained from the quotients closely as in this case. On the other hand, the evidence that this is true is sufficient to carry considerable weight and to justify one in proceeding tentatively on this basis until further data bearing on the point have been collected and published.

The second of the methods which the writer found to give approximations to the actual probable errors is that used in the case of sums and differences. The standard formula for the probable error of a sum or difference on the assumption that the quantities composing it are uncorrelated is as follows[3]: $\text{P.E.}_{x \pm y} = \sqrt{\text{P.E.}_x^2 + \text{P.E.}_y^2}$. It occurred to the writer that the probable error of a quotient might be of somewhat the same size as that of a difference; therefore he experimented with this formula. If the probable errors of measurement of the achievement and mental ages are substituted therein, the results are not at all similar to those actually obtained for achievement quotients. The reason for this is easy to see; the ages are expressed in terms of an entirely different unit from that used in the quotients. In the case of the formula discussed above, this situation was taken care of by dividing by the denominator, but in the formula just given, no such division occurs, or, in other words, nothing is done to change the unit employed from that of the numerator and the denominator to that used in the quotient. This lack of the same units may be taken care of, however, by dividing the probable errors of measurement by their means or standard deviations. The writer, therefore, tried out the formula for the probable error of the difference by substituting in it the ratios of the probable errors of the achievement and mental ages, respectively, to their means and standard deviations. The results along with the actually obtained similar ratios previously given in Table V are shown in Table VII. The first pair of columns in the table compares the results by the two methods for the ratios of the probable errors of measurement to the means and the

[3]Adapted from:
Yule, G. U. *An Introduction to the Theory of Statistics.* London: Charles Griffin and Company, 1919, p. 211.

TABLE VII. COMPARISON OF RATIOS OF PROBABLE ERRORS OF MEASUREMENT TO THE MEAN AND STANDARD DEVIATION COMPUTED BY THE FORMULA FOR THE PROBABLE ERROR OF A DIFFERENCE WITH THOSE OBTAINED FROM THE ACTUAL ACHIEVEMENT QUOTIENTS

	$\frac{P.E._{meas.}}{M}$		$\frac{P.E._{meas.}}{\sigma}$	
	By Formula	From Quotients	By Formula	From Quotients
Stanford Arithmetic	.05	.05	.44	.45
Monroe Arithmetic	.08	.08	.50	.47
Monroe Reading Comprehension	.09	.09	.53	.53
Monroe Reading Rate	.10	.11	.50	.44

second pair of columns for the ratios to the standard deviations. As will be seen, the entries in the first pair carried to the second decimal place differ in only one case out of the four, and this difference is as small as possible, being only .01. The differences in the case of the entries in the second pair of columns are somewhat greater, though not very large. On the whole it appears that the formula for the error of a difference gives an approximately correct error for a quotient also when the ratios of probable errors to means are employed, and a somewhat less satisfactory one when their ratios to the standard deviations are used. Probably it should not be applied in connection with the latter, although the errors involved in doing so are not great. There is little choice between the two methods of approximation, the first being slightly more accurate for the ratio of the probable error to the mean, and the second for its ratio to the standard deviation.

APPENDIX C

A COMPARISON OF THE RELIABILITY OF QUOTIENT MEASURES AND DIFFERENCE MEASURES

From the standpoint of reliability, there appears to be little choice between difference measures and quotient measures. Foran[1] is among those making the statement that there is little difference between the two kinds of measures with regard to reliability, but neither he nor anyone else, in so far as the writer knows, has cited suitable data to prove this contention. From a logical standpoint, it seems reasonable that if measures of achievement and of capacity possess given amounts of unreliability which tend to be cumulative for measures of achievement relative to capacity, the total unreliability would be about the same for either difference or quotient measures. In order to investigate the truth of this assumption, however, the writer found the probable errors of measurement for both difference and quotient measures computed from the same test scores. The results are given in Table VIII. The first half of this table contains the coefficients of correlation and of alienation, the probable errors of measurement and the ratios of these errors to the means and to the standard deviations for differences between achievement scores and intelligence scores. The second half of the table contains the same items for the quotient measures derived from the same test scores as furnished the basis for the difference measures.[2]

It will be seen from the figures in this table that there is a slight tendency for the coefficients of correlation of the difference measures to be larger and the corresponding coefficients of alienation smaller, than those of the quotient measures, but that this tendency is not strong enough to be significant. The probable errors of measurement of the difference measures are distinctly greater than those of the quotient measures, but their ratios to the means and standard deviations are in most cases slightly less than the corresponding ratios for the quotient measures. On the whole, although these data show that the reliability of the difference measures is greater than that of the quotient measures, the difference is so small as to justify the assumption that there is little difference in the reliability of difference and quotient scores.

[1]Foran, T. G. "The Meaning and Limitations of Scores, Norms, and Standards in Educational Measurement," *Catholic University of America Educational Research Bulletin,* Vol. 3, No. 2. Washington: Catholic Education Press, February, 1928, p. 16-19, 23-26.

[2]The material in the second half of the table is taken from Table V on p. 45.

TABLE VIII. COMPARISON OF RELIABILITY OF DIFFERENCE MEASURES AND QUOTIENT MEASURES DERIVED FROM THE SAME TEST SCORES

	Difference Measures					Quotient Measures				
	r	k	$P.E._{meas.}$	$\frac{P.E._{meas.}}{M}$	$\frac{P.E._{meas.}}{\sigma}$	r	k	$P.E._{meas.}$	$\frac{P.E._{meas.}}{M}$	$\frac{P.E._{meas.}}{\sigma}$
Stanford Arithmetic	.57	.82	5.7	.04	.44	.54	.84	4.6	.05	.45
Monroe Arithmetic	.54	.84	16.9	.07	.46	.53	.85	9.6	.08	.47
Monroe Reading Comprehension	.34	.94	17.1	.09	.55	.39	.93	9.2	.09	.53
Monroe Reading Rate	.64	.77	21.8	.10	.40	.58	.81	13.0	.11	.44